Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 12 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Review

Learning and Use of eHealth Among Older Adults Living at Home in Rural and Nonrural Settings: Systematic Review

Ella Airola[1], MEd

Media Education Hub, Faculty of Education, University of Lapland, Rovaniemi, Finland

Corresponding Author:
Ella Airola, MEd
Media Education Hub
Faculty of Education
University of Lapland
P.O. Box 122
Rovaniemi, FI-96101
Finland
Phone: 358 44 4744389
Email: ella.airola@ulapland.fi

Abstract

Background: Care policies emphasize deinstitutionalization and *aging in place* in response to demographic changes. Different eHealth technologies are one way to achieve this aim. However, there is a need to better understand older adults' needs for eHealth services, and thus, these health solutions require further exploration.

Objective: The purpose of this systematic literature review is to appraise, synthesize, and summarize the literature on older adults' (aged ≥60 years) eHealth learning and use in real home settings, particularly in rural and remote areas, with a focus on the social and cultural context.

Methods: A systematic search was conducted in January 2020 using 4 academic databases. The studies by means of qualitative thematic analysis to identify the barriers, enablers, and support practices involved in the domestication process were examined. In addition, we identified the various meanings attached to eHealth technologies for older adults living in rural and remote areas.

Results: In total, 31 empirical studies published between 2010 and 2020 were included in this review. A total of 17 articles included participants from rural and remote areas. The most regularly reported barriers related to older adults' learning to use and use of eHealth were health-related difficulties, such as cognitive impairment or impaired hearing. The most reported enabler was the support provided for older adults in learning and use of eHealth. Support mainly comprised older adults' own digital competences, which were distributed with their social network. It was found that eHealth technology is needed for rural and remote areas to facilitate access and reduce logistical barriers to health care services.

Conclusions: The literature review provided information and practical implications for designers, health care providers, and policy makers. On the basis of these findings, eHealth technologies should be easy to use, and adequate support should be provided to older adults for use.

(J Med Internet Res 2021;23(12):e23804) doi:10.2196/23804

KEYWORDS

aged; barrier; digital competence; deinstitutionalization; eHealth; home care; learning; older adult; rural health

Introduction

Background

The aim of this systematic literature review is to advance the understanding of older adults' (aged ≥60 years) eHealth learning and use during its domestication. The main target is older adults living in rural and remote environments, as they are far removed from traditional health care services [1-4]. According to the World Health Organization, health is understood as "a state of complete physical, mental, and social well-being and not merely the absence of disease of infirmity" [5]. As a general practice, eHealth can be seen as an umbrella concept for different health care services delivered or improved through information and communication technologies (ICTs) [6,7].

On the basis of future demographic changes, the world will face two challenges: a growing care burden per capita and

organizational changes to health care systems [8,9]. To achieve its sustainable development goals, the United Nations [8] has outlined key policy changes that deal with current and future population dynamics. Recommendations include investing in life-long learning, especially digital competences, and promoting healthy aging and long-term care systems to meet the needs of the aging population. Care policy throughout Europe has emphasized deinstitutionalization and *aging in place* to reduce paternalistic care and improve the quality of care [10,11]. Furthermore, previous research in many cultural settings has found that older adults prefer living at home for as long as possible [9].

Theoretical Framework

This study's theoretical framework is built on the concept of *distributed and situated digital competence*, *technology domestication*, and previous research on *barriers and enablers* of technology use. The European Commission's *DigComp 2.0: The Digital Competence Framework for Citizens* [12] defines digital competence as a combination of knowledge, skills, and attitudes related to the use of ICT tools. One competence is the protection of health and well-being, which means, for example, being aware of digital technologies that can be used to enhance social well-being and social inclusion. However, the framework can be criticized for being decontextualized and centered around individuals. Therefore, a socially and culturally oriented approach is followed, and digital competences are understood as the distributed and situated competences of older adults and their social networks [13,14].

The concept of domestication focuses on how technology users and nonusers adopt technologies culturally and socially in their everyday lives [15-17]. The domestication of technology includes four dimensions: appropriation, objectification, incorporation, and conversion. The process starts with appropriation, which does or does not create a relationship with the new technology, and continues to objectification, which is when the technology is given a place at home. Incorporation focuses on the technology's place in and influence over the user's everyday routines. Finally, the technology becomes familiar, and conversion is achieved [15-21]. Central to domestication are the public and personal meanings attached to the technology, which actively transform during the process [19]. Meanings are understood as individuals' thoughts and feelings regarding the technology and how it is seen as part of his or her own cultural context [22,23].

Research into health-related barriers has found that the key reasons older adults avoid ICT adoption are the accessibility of support, their health status, lack of need or interest, functionality, added value, cost, and concerns regarding privacy and trust [24,25]. In the present review, barriers were not understood as insurmountable obstacles but rather as challenges that can be solved. Conversely, the following key enablers influence older adults' use of eHealth: motivation, support, and feedback [26]. In addition, meaningful function and aesthetics are important when older adults adopt technology [27].

Objective

Recent international literature reviews on eHealth and older adults focus on a variety of topics: eHealth access and use from a health equity perspective [28], the facilitators of and barriers to eHealth use [24,26,29], and user involvement in technology design, including eHealth technologies [30,31]. In addition, Cheng et al [32] published a systematic review of eHealth interventions targeted at socially disadvantaged groups, including older adults. However, we still lack a review that examines older adults' eHealth learning and use in real home settings and in a rural context from users' perspectives. The purpose of the present review is to fill this gap by focusing on studies that report barriers and enablers that older adults face when learning to use and while using eHealth technology and how older adults are supported during the process. In addition, what eHealth technology means to older adults in rural and remote areas was examined. In reviewing relevant studies, this research seeks to answer the following questions:

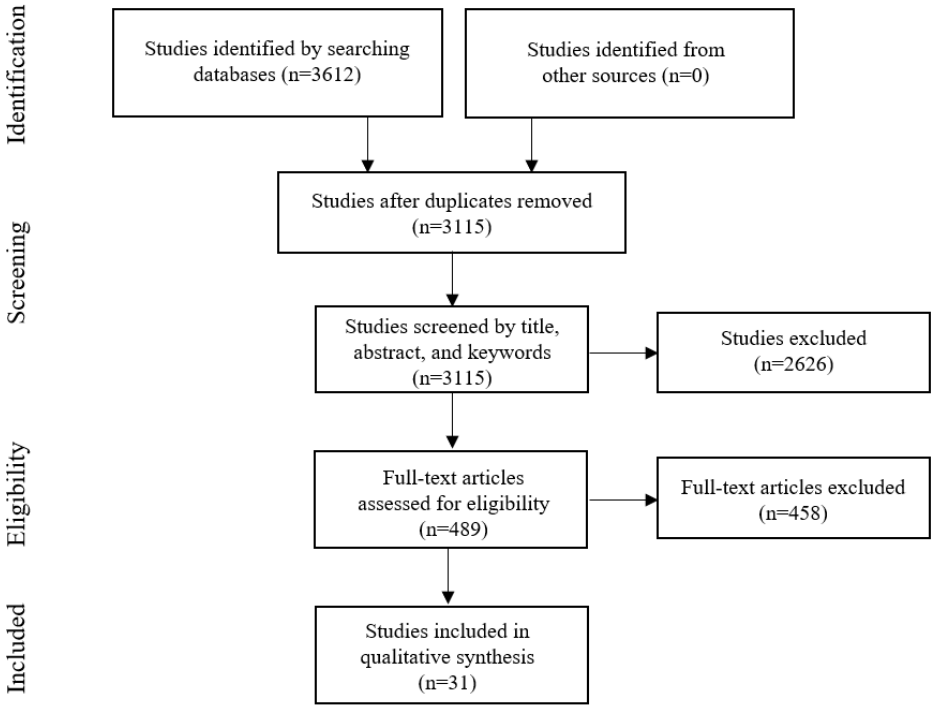
1. What barriers and enablers are related to the learning and use of eHealth technologies in domestication processes among older adults living at home?
2. How are older adults living at home supported in their domestication of eHealth technologies?
3. What are the meanings attached to eHealth technologies for older adults living in rural and remote areas?

Methods

Overview

The methodology follows the standard PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for studies that evaluate health care research [33]. PRISMA includes a 27-item checklist (Multimedia Appendix 1 [33]) and a 4-phase diagram (Figure 1) to help authors ensure clarity and transparency. A systematic literature review was selected "to appraise, synthesize, and summarize" [34] the literature and determine whether further primary research in the area is needed. The entire review process was undertaken by 1 author, and there were no additional authors.

Figure 1. Flow of information through the different phases of the systematic review.



Search Strategy

To begin the search process, the author consulted an information specialist at her university library regarding search phrases and relevant databases. A literature search was conducted in January 2020 and included the following international web-based databases: ProQuest (ERIC—Education Collection, Social Science Database, Applied Social Sciences Index and Abstracts, Sociological Abstracts, and Sociology Database), Ebsco (AgeLine, Academic Search Elite, and CINAHL Complete), Web of Science (Arts and Humanities Citation Index, Social Sciences Citation Index, Science Citation Index Expanded, and Emerging Sources Citation Index), and Scopus (Elsevier). Relevant articles were found using the search terms specified in Textbox 1 and their combinations. An example search string used to search Scopus (Elsevier) can be found in Multimedia Appendix 2. The search was applied to the field title, abstract, and keywords. Although the conversions around eHealth began at the turn of the millennium [6], the review's start date of 2010 was chosen as there has been a lot of change in health technologies over the past decade [35]. The initial database search yielded 3612 papers (n=393, 10.88% from ProQuest, n=394, 10.91% from Ebsco, n=2767, 76.61% from Web of Science, and n=58, 1.61% from Scopus). No articles were included that were identified from sources other than the abovementioned databases.

Textbox 1. Content and search terms of literature search. During the search, some terms were truncated using "*" so the search considers all the results including the first part of the word. For example, by searching learn*, the search also considers terms learner and learning.

Content and search terms

- Older people: *older, senior, elderly, aged people, old age user, elder*
- Rural: *rural, remote, sparsely populated area*
- Use: *education, learn*, competence, digital skill, geragogy, use, reject, active aging, adoption, acceptance, barrier, enabler, facilitator*
- Digital technology: *online, ICT, information, computer, internet, electronic, techolog*, digital, smart, management tool, virtual, mobile, robot, tele*, monitoring, assist*, gerontechology, compliance, reminder, dispens*, video, application, device*
- Health: *health, care, wellbeing, physical, mental, social*
- Home: *aging in place, independent, home, everyday life, living, daily life, domestication*

Study Selection

Figure 1 shows a flowchart depicting how PRISMA was used to select the studies. All articles were exported to Refworks (ProQuest), a reference management system that automatically discards duplicate papers. Of the 3612 papers, the removal reduced the number of papers to 3115 (86.24%). The remaining articles were screened for title, abstract, and keywords to select studies relevant to the research topic. Following the screening of the 3115 articles, a total of 489 (15.7%) articles were obtained. The remaining articles were exported to NVivo 12 (QSR International), a qualitative data analysis software, and the remaining duplicates were manually discarded. Full-text

articles were then assessed for eligibility based on the preliminary inclusion and exclusion criteria.

Finally, full-text articles (31/489, 6.3%) that met the following criteria (Textbox 2) were accepted: published in a peer-reviewed scientific journal; written in English; published between January 1, 2010, and January 22, 2020; had an empirical study design aimed at supporting older adults' use of eHealth technology in real home settings at the user's home; completely or partially focused on technology users' perspectives; had participants with a mean age of 60 years; and included participants who were older adults with or without health conditions. Initially, the aim was to select studies that included only older adults aged >65 years, which is a common age limit for older adults in research [36]; however, because of the low number of studies, the age range was extended to a mean of 60 years. In addition, living in rural or remote areas was a planned inclusion criterion; however, it was modified to living in any area, as not enough studies met the original criterion. However, one research question focused only on studies conducted in rural and remote areas.

Textbox 2. Inclusion and exclusion criteria.

Inclusion criteria

- Academic paper published in a peer-reviewed scientific journal
- Written in English
- Published between 2010 and 2020
- Had an empirical study design aimed at supporting older adults' eHealth use
- Was located in real home settings at the user's home
- Had participants with a mean age of 60 years
- Completely or partially focused on users' perspectives

Exclusion criteria

- Did not include older adult participants
- Was not conducted in real home settings
- Was conducted in a laboratory or included no user experience with eHealth
- Was a review or theoretical study
- Only had a biomedical or technical perspective

Data Exclusion and Analysis

For the final set of full-text articles, a data extraction sheet was developed (Multimedia Appendix 3 [37-65]). The gathered data included information on article titles, authors, journals, publication years, methods, eHealth technologies, target groups, total number of participants, participant genders and ages, participants' area of residence (eg, rural or remote), and countries of residence. The included articles were analyzed in NVivo 12 using a qualitative thematic approach guided by the concepts of domestication and digital competence. In addition, previous research on the barriers to and enablers of technology use was used. The author thoroughly read and reread the articles and identified 154 subcategories that described significant sentences and phrases concerning the research questions. The obtained subcategories were clustered into 8 upper-level categories to produce a more nuanced understanding.

Results

Study Characteristics

The selected 31 articles were published in 29 different journals representing multidisciplinary research. Of the 31 papers, 16 (52%) were quantitative, 4 (13%) were qualitative, and 11 (35%) were mixed methods studies. Most studies used >1 method; only 16% (5/31) used just one method. The most common qualitative data collection methods were surveys and questionnaires (20/31, 65%) and technical logs (12/31, 39%). The most common qualitative data collection methods were interviews (9/31, 29%) and focus groups (5/31, 16%). The studies were conducted in the following 12 countries: the United States (9/31, 29%), Canada (2/31, 6%), Italy (2/31, 6%), Spain (2/31, 6%), Sweden (2/31, 6%), the United Kingdom (2/31, 6%), France (1/31, 3%), Germany (1/31, 3%), Ireland (1/31, 3%), Lebanon (1/31, 3%), Norway (1/31, 3%), and Switzerland (1/31, 3%). In 19% (6/31) of studies, the country was not identified.

The number of participants ranged from 1 to 4380, although not all were older adults. In all articles, either the mean or median age and SD or the age range was defined. However, in studies in which only the SD and mean or median age were given, it was not possible to strictly count the age of the youngest and oldest participants. In 90% (28/31) of articles, the mean or median age of the technology users was ≥60 years, and in 39% (12/31) of articles, the minimum age was >60 years. Some articles (6/31, 19%) included technology users aged<50 years, and it was possible to separate them from the older adults in the findings. This means that the findings were still based on participants who were aged >60 years or whose mean or median age was ≥60 years. In 42% (13/31) of articles, most participants were women, and in 35% (11/31) of articles, most participants

were men. In 6% (2/31) of papers, female and male participants were equally represented, and in 16% (5/31) of papers, gender was not defined.

All participants used eHealth technologies in real home settings. The eHealth technologies that occurred most frequently were mobile apps (18/31, 58%), monitoring systems (12/31, 39%), web-based platforms (11/31, 35%), assistive technologies (2/31, 6%), and ambient awareness technology (1/31, 3%). No eHealth robots were included in this study. Most technologies were targeted at older adults with different noncommunicable diseases. Of those, the largest target groups were cardiac disease (5/31, 16%), diabetes (3/31, 10%), and cancer (3/31, 10%). In total, 2 technologies were used for older adults with polypharmacy issues and 2 for those in palliative care. In 16% (5/31) of studies, eHealth technologies were aimed at isolated older adults, and in 13% (4/31) of studies, the technology was for older adults with depression, anxiety, or apathetic qualities. In total, 1 eHealth technology was only for healthy older adults, and 5 were for older adults or adults of any age group in general. In some studies, the participants were veterans (4/31, 13%), and in others, they were older family caregivers (4/31, 13%).

Barriers and Enablers in the Domestication Process

Overview

The first research question focused on the barriers and enablers related to the learning and use of eHealth technologies in the domestication process among older adults living at home. Thematic analysis produced 111 subcategories related to these barriers and enablers. The subcategories were clustered into 4 dimensions of the domestication process [15,17]: appropriation (40/111, 36%), objectification (10/111, 9%), incorporation (27/111, 24.3%), and conversion (34/111, 30.6%). A summary of these findings is presented in Tables 1 and 2.

Table 1. Summary of the findings concerning the barriers to learning and use of eHealth technologies among older adults (N=31).

Dimension of domestication and barriers	Papers, n (%)
Appropriation	
eHealth technology	
Lack of connectivity	9 (29)
Technical problems	8 (26)
Difficult to use	6 (19)
Unclear instructions	5 (16)
Cost	4 (13)
Technical limitations	4 (13)
Difficult to learn to use	3 (10)
No feedback	2 (6)
Lack of effectiveness	1 (3)
Lack of technical device	1 (3)
Technology was unexpected	1 (3)
Old age user	
Health-related difficulties	17 (55)
Lack of previous experience	5 (16)
Uncertain with the technology	5 (16)
Irritation or frustration	4 (13)
Lack of motivation or interest	2 (6)
Personal factors	2 (6)
Fatigue	2 (6)
Being intimidated	1 (3)
Lack of digital competence	1 (3)
Skeptical	1 (3)
Unwilling to use the technology	1 (3)
Objectification	
Object at home	
Design	2 (6)
Placement in the home	2 (6)
Ergonomics	1 (3)
Data protection and security of the eHealth	
Concerns about security or privacy	4 (13)
Lack of reliability	2 (6)
Incorporation	
Everyday life	
Unsuitable for everyday life	3 (10)
Time constraints or power dynamics	3 (10)
Lack of utility	2 (6)
Inappropriate technology	1 (3)
Logistical difficulties	1 (3)
Not meaningful information	1 (3)
Not meaningful service	1 (3)

Dimension of domestication and barriers	Papers, n (%)
Technology used only occasionally	1 (3)
Conversion	
Social interactions	
Need for face-to-face contact	7 (23)
Lack of support	3 (10)
Family relationships	2 (6)
Feeling like an outsider	1 (3)
Lack of communication	1 (3)
Lack of patient–professional communication	1 (3)
Shyness	1 (3)
unable to use independently	1 (3)
Society and culture	
Older age	4 (13)
Not culturally relevant	3 (10)
Living in rural area	2 (6)
Lower socioeconomic status	1 (3)
Being female	1 (3)

Table 2. Summary of the findings concerning the enablers of learning and use of eHealth technologies among older adults (N=31).

Dimension of domestication and enablers	Papers, n (%)
Appropriation	
eHealth technology	
Usability	18 (58)
Personalization or flexibility	13 (42)
Familiarity	9 (29)
Feedback	7 (23)
Accessibility	3 (10)
Automated service	1 (3)
Novelty effect	1 (3)
Offers personal challenge	1 (3)
Positive experiences of others	1 (3)
Old age user	
Satisfaction	14 (15)
Confidence or self-esteem	4 (13)
Self-efficacy	4 (13)
Feeling of success	3 (10)
Open-minded	3 (10)
Digital competence of the user	2 (6)
Interest in electronic devices	1 (3)
No feeling of privacy loss	1 (3)
Own choice	1 (3)
Objectification	
Object at home	
Placement in the home	2 (6)
Design	1 (3)
Data protection and security of the eHealth	
Technology's security or safety	2 (6)
User privacy	2 (6)
Reliability	1 (3)
Incorporation	
Everyday life	
Suitable for everyday life	10 (32)
Service provided educational information	8 (26)
Active part of user's own care, self-care	7 (23)
Playful	7 (23)
No logistical barriers	7 (23)
No temporal barriers	6 (19)
Improve quality of daily life	6 (19)
Useful service	6 (19)
Increased security	5 (16)
Systematic use of the technology	4 (13)
Extends one's own habitat	3 (10)

Dimension of domestication and enablers	Papers, n (%)
Healthier lifestyle	3 (10)
Meaningful service	3 (10)
No financial barriers	3 (10)
Brings joy	2 (6)
Fewer clinical visits	2 (6)
Technology provides freedom	2 (6)
Helpful	1 (3)
Not in a hurry	1 (3)
Conversion	
Social interactions	
Support practices	27 (87)
Social connectedness and belonging	10 (32)
Patient–professional communication	9 (29)
Visual contact	4 (13)
Shared experience	3 (10)
Face-to-face meeting	2 (6)
Group assignments	2 (6)
Social comparison	2 (6)
Feeling less lonely	1 (3)
Individual attention	1 (3)
Loneliness	1 (3)
Relationship with technology	1 (3)
Society and culture	
Cost-effectiveness	8 (26)
Cultural relevance	4 (13)
Supports independence	2 (6)
Feeling of being equal	2 (6)
Being female	2 (6)
User feels like an active citizen	1 (3)
Being considered *cool*	1 (3)
Educated or employed	1 (3)
Older age	1 (3)

Appropriation

Of the 31 articles, 30 (97%) reported barriers and enablers in the appropriation phase of the domestication process; this dimension had more barriers than the others. Older users' *health-related difficulties* was the most common barrier in the appropriation phase and the whole domestication process (17/31, 55%). For example, difficulties included cognitive impairment or dementia [37-39] and impaired hearing or vision [40,41]. The most common barriers related to the eHealth technologies themselves were *a lack of connectivity* (eg, a lack of access to fast enough internet connections), which was reported in 29% (9/31) of articles [42,43] and other *technical problems*, which were reported in 26% (8/31) of articles [44,45]. *The usability* of eHealth technology, which included ease of use, was the ruling enabler and was reported in 58% (18/31) of studies [46,47]. *Satisfaction* was also seen as a significant enabler in the domestication process and was reported in 45% (14/31) of studies [53,66]. In addition, *personalization or flexibility* of eHealth technology was reported as an enabler in 42% (13/31) of studies [48,49]. That the technology o*ffered a personal challenge* was reported as an enabler in 3% (1/31) of article [37]; the article reported that if the tasks proposed in a mental training system are slightly above users' capabilities, users' motivation is increased and training outcomes are improved. That the *technology was unexpected* was reported as a barrier in 3% (1/31) of articles [51], and this refers to users feeling they

had received a remote monitoring system unexpectedly without a choice and without knowing what the system was.

Objectification

Objectification was the least-coded dimension, and there was not a big difference between the different subcategories. The *design* of the eHealth technology and its placement in the home were reported as barriers and enablers. As a barrier, design referred to the look of a computer [50] and the weight and size of a remote monitoring system [51]. As an enabler, it was connected to the range of soundscapes provided to the users in an interaction radio [52]. In 3% (1/31) of articles [50], it was found that if a technology is placed *out of the way*, the placement can be a barrier to its use; however, if it is placed where the user spends time, it is used more often. In the Ottenberg et al [51] study, in addition to the weight and size of the system, wiring configurations challenged its placement. Furthermore, *ergonomic challenges*, such as typing or using a mouse, were reported as barriers in 3% (1/31) of articles [40], and this was in the context of a web-based pain management program designed for older adults with chronic pain. The objectification dimension also included barriers and enablers related to data protection and security. Here, the most common barriers were *concerns about security or privacy* [39,51,53,54]. The leading enablers were *technology security or safety* [38,51] and *user privacy* [45,50].

Incorporation

Of the 31 articles, 24 (77%) reported barriers and enablers during the incorporation phase. As reported in 32% (10/31) of articles, the most common enabler was that the eHealth technology was *suitable for everyday life* of older adults. For example, this was evident in how users did not actively think they used the technology [52], were able to use it in a self-selected environment [38,53], and found that it matched their needs [55,56]. A total of 2 equally common barriers, each reported in 10% (3/31) of articles, were that the technology was *unsuitable for everyday life* and that there were *time constraints or power dynamics* in using it, which means, for example, that users had scheduling conflicts [38], were busy [41,53], or expected faster replies via the technology [53]. The reasons the technology was not suitable for users' daily lives were that the users had to change or constrain their behaviors when using it [52], were busy or on holiday [53], and had experienced a stressful life event that caused them to disengage with the service [56].

Conversion

The conversion dimension had more enablers than the other dimensions. The most common enabler in the conversion phase and the entire domestication process was *support practices*, which was reported in 87% (27/31) of articles. The findings are presented in detail in *the Support Practices During Domestication* section below. Regarding social interactions, *social connectedness and belonging* were mentioned as enablers in the learning and use of eHealth technology in 32% (10/31) of studies [44,57]. The common barriers related to social interactions were that older adults had a *need for face-to-face contact* [39,51] and felt a *lack of support* while using the technology [44,51,58].

Regarding society and culture, *cost-effectiveness* was the most common enabler in the conversion phase, as reported in 26% (8/31) of articles. However, it was almost never reported from a user's perspective and was more society driven, providing "huge savings in clinical costs" [59], "significant saves in the health budgets" [60], and "suggesting a cost reduction for the health care system" [46]. In addition, in 13% (4/31) of articles, *cultural relevance* was noted as an enabler, which means, for example, that the users could incorporate their religious or spiritual perspectives into the service [48] or that televisions were used as devices, as they were best integrated into the older adults' lives [37,59].

In addition, o*lder age* and *being female* were reported as enablers and barriers in different articles. For example, 3% (1/31) of articles [60] reported that women needed more assistance using a mobile health app than men, and in 6% (2/31) of articles [53,67], women were more active eHealth technology users than men. Older age was mainly seen as a barrier in cases where older adults needed more help using the technology [60] or used it less than younger people [60-62,67]. However, in 3% (1/31) of articles [60], older patients reported greater improvements in several measures than younger patients. In 3% (1/31) of articles [62], it was reported that *a lower social class* reduced the prevalence of eHealth technology use.

Support Practices During Domestication

To understand how older adults were supported while domesticating eHealth technologies, the thematic analysis produced 25 subcategories related to their learning to use and the use of technology. The subcategories were clustered into two upper-level categories: social network support (22/31, 71%) and nonsocial support (3/31, 10%). Of the 31 articles, 27 (87%) reported that support was provided to users during the domestication process [15,17].

Social Network Support

Social network support was provided in 87% (27/31) of articles. During the learning process at the beginning of domestication, *face-to-face support* was more common than long-distance support. Face-to-face support was provided to users in 45% (14/31) of articles and long-distance support in 13% (4/31) of articles. The most common way to teach older adults was *a training session*, which was used in 42% (13/31) of articles. Training sessions were used to introduce the users to the use of the technology, and it could happen either at users' homes [45,47] or in a clinical setting [38,47]. The training was usually given by technical staff [59], health care professionals [49,63], or the researcher of the study [53]. *Long-distance support* was more common than face-to-face support after initial training. Long-distance support was provided to users in 39% (12/31) of articles and face-to-face support in 10% (3/31) of articles. Long-distance support was mostly provided *via phone*, and it was mainly related to technical issues [59] or questions related to the study [61], although it also took the form of consultation with a health care professional [42] or communication with a peer [38]. In some articles, it was not possible to identify

whether social network support was face-to-face or long-distance.

During the domestication process, eHealth technology users distributed their digital competence [13,14] with the following members of social networks: *spouses, grandchildren, other family members and relatives, family caregivers, peers, friends, health care professionals, coaches, researchers, technical staff,* and *proctors*. Social networks expanded as the domestication process went further. Digital competences were most commonly distributed among health care professionals, such as nurses, therapists, and social workers. This was reported in 48% (15/31) of articles [42,44]. Second, technical staff was mentioned in 23% (7/31) of studies [39,63].

Nonsocial Support

Nonsocial support practices in the learning process were *written instructions* [37,47,49,57,63], *video instructions* [53], and *diaries* [52]. Users owned their diaries before engaging in the study, and their use to support learning was not planned beforehand. Nonsocial support practices were not defined after users passed the introduction phase; however, in 3% (1/31) of articles, users were encouraged to "access the video at any time during the trial" [53]. None of the articles reported the provision of only nonsocial support. Nonsocial support was always connected to social network support.

Meanings of eHealth in Rural and Remote Areas

Overview

The focus of the third research question was on the meanings attached to eHealth technologies for older adults living in rural and remote areas. The terms *rural*, *remote*, and *sparsely populated area* were used to search the databases (Textbox 1) for this literature review. However, in only 55% (17/31) of articles, every participant (12/17, 71%) or part of the participants (5/17, 29%) lived in rural or remote areas. The minimum amount of rural or remote participants per selected article was 13.29%. The thematic analysis produced 17 subcategories of meanings attached to eHealth technologies for older adults living in rural and remote areas. The subcategories were clustered into 2 upper-level categories: needed for rural and remote areas (13/17, 76%) and source of inconvenience and concern (4/17, 24%). The articles defined rural and remote areas as underserved areas with limited access to health care.

Needed for Rural and Remote Area

Of the 17 articles, 14 (82%) reported that different eHealth technologies, such as home telehealth monitoring and videoconferencing for consultation, were needed in rural and remote areas. A total of 2 meanings clearly stood out: first, 71% (12/17) of articles [40,56] reported that eHealth technologies *facilitate access to health care services* for older rural adults; second, 65% (11/17) of articles [48,64] reported that there are *no logistical barriers* to health care services when older rural adults use eHealth technology at home. Other, less commonly reported meanings were *no temporal barriers* [61], *no financial barriers* [48], *supports relationship with care provider* [64], *no physical or physiological stress* [39], *ability to support rural caregivers* [42], *reduces boundaries of home* [50], *reduces feelings of isolation* [44], *ensures equal access to health care services* [41], *increases feelings of security* [44], *no weather-related barriers* [39], and *permits religious or spiritual inclusion* [48].

Source of Inconvenience and Concern

Although eHealth technology was principally seen as needed for rural and remote areas, 24% (4/17) of articles identified sources of inconvenience and concern. Here, the most commonly coded meanings are related to internet connectivity and use. Approximately 18% (3/17) of studies reported that rural areas *lack access to high-speed internet* [40,42,61]. Approximately 12% (2/17) of studies [42,61] noted *cultural differences regarding internet use* in rural areas compared with urban areas. Older adults in rural areas are still uncomfortable using the internet. Approximately 6% (1/17) of studies [45] reported that living in rural areas required *additional equipment*, such as "protectors to protect equipment," and rural participants had to *learn how to reset* the technology "to decrease the need for providers to make a home visit specific to technical support." The latter was the only meaning assigned to learning to use eHealth technology in rural areas.

Discussion

Principal Findings

To advance the understanding of older adults' (aged ≥60 years) eHealth learning and use during its domestication, a systematic literature review of 31 empirical studies published between 2010 and January 2020 was conducted. The aim was to summarize the literature on the barriers and enablers that older adults encounter when learning to use and using eHealth technology and how they are supported in real home settings. The main targets were rural and remote older adults. The key findings of this review confirmed that social networks supporting older adults are important enablers for learning how to use and using eHealth technology. In addition, this review revealed that health-related difficulties often prevent older adults from domesticating eHealth technologies. Various eHealth technologies have been reported as necessary for older adults in rural and remote areas, although some sources of inconvenience and concern related to internet connectivity and use were found.

One of the goals of this review was to find out which barriers and enablers are related to learning how to use and using eHealth technologies in the domestication processes among older adults living at home. The findings were divided into 4 dimensions (appropriation, objectification, incorporation, and conversion) of the domestication process [15,17]. The barriers and enablers are in line with previous similar systematic reviews [24,26]. The accessibility of support and training, usability of the technology, and its suitability for daily life were also stressed in the present review. However, for example, concerns regarding the cost or importance of motivation to use eHealth technologies were not key factors in this review. There was a significant difference between the number of reported barriers (n=48) and enablers (n=63), with enablers being more common. This may be because of a desire to highlight the benefits of eHealth technology in a study, even if unintentionally. The appropriation

dimension had the most barriers, which suggests that the first phase of domestication is critical. Previous research [68,69] has expressed that access and support for older adults using technology, especially in the early weeks of its domestication, may be the most important factor in successful technology adoption. Overall, the studies focused less on the learning process than on the use of technology.

In addition, this systematic review asked how older adults living at home are supported during the domestication of eHealth technology. In most cases, older adults distributed their digital competences with their social networks, which supported the domestication of eHealth technology. The social network included *warm experts* who are "nonprofessional persons who help inexperienced users come to terms with digital devices" [70], such as family members and peers. The network also included formal personnel, such as health care professionals and technical staff. Previous qualitative research related to eHealth learning and use [71] has underlined that there is a need to clarify the role of peer-to-peer support in the domestication of technology. This study revealed that peer-to-peer support has its place in the digital health technology context [38,44,52] as well as in other settings [14]. However, health care professionals were found to be the ones with whom digital competences were most often distributed; therefore, this review argues that digital competences related to eHealth learning and use are not only technical skills but also knowledge and skills related to health care, which is why health care professionals are also needed to support eHealth technology users.

This review's final aim was to better understand the meanings attached to eHealth technologies for older adults living in rural and remote areas. Although the literature includes few studies set in rural and remote areas, this review confirms that eHealth technologies are needed in rural and remote areas for several reasons. eHealth technology is seen as solving many problems related to limited access to health care services, such as logistical, temporal, financial, and weather-related barriers. In addition, eHealth technologies can foster a sense of belonging in older rural residents by reducing feelings of isolation and by connecting them with peers or care providers. Previous research has also shown that ICT services can reduce social isolation and promote social connectivity in older adults experiencing physical and cognitive decline or living in remote areas [3,72,73]. However, the use of eHealth technology in rural and remote areas is hampered by the lack of high-speed internet connections and older adults' lack of comfort in using ICTs. These issues are targets for future development. The findings of this systematic review confirm that the use and nonuse of eHealth technology are related to its fit in older adults' technology-related cultural understanding and the context in which people act and live [13,14].

This review reveals several gaps that suggest directions for future research. First, a stronger focus on older adults' learning processes is required in research on the domestication of eHealth technology. In the literature search, numerous articles focused only on older adults and the use of eHealth technology from a biomedical [74] or technical perspective [75]. They were excluded from the review as they did not focus on the user's point of view. Burholt and Dobbs [2] also found that research on rural aging is dominated by a biomedical perspective. Second, as this review confirmed, eHealth technology is needed in rural and remote areas. Therefore, more studies in these settings are required to better understand the processes of domesticating eHealth technology among older rural residents. Third, research efforts in eHealth technology studies in real home settings are limited. Several studies were excluded from this review as they focused on a communal dwelling [76] or nursing home [77], although the aim of the care policies is aging in place [10], which may be because it is easier to reach older adults and conduct research in organized environments. Finally, methodologically, studies on the domestication of technology were mostly qualitative, and mixed methods studies were most common [18,78]. However, the studies reviewed here were mainly quantitative. Therefore, future qualitative studies on this topic should include the voices of older adults. Furthermore, by the time this review is published, new papers relevant to this topic will have also been published. Therefore, new systematic literature reviews should be regularly conducted on the topic of eHealth knowledge and use among older adults living at home in rural and remote areas.

Limitations

This review has several limitations. First, as there was only 1 reviewer involved, this review's quality and reliability were somewhat weakened. Additional reviewers would have challenged the decisions made by the author. Second, although the selected studies were conducted in 12 countries, Western countries and English-speaking countries dominated, and studies not written in English were excluded. As noted in earlier reviews [30], this may create a bias by, for example, basing the findings mainly on reports from Western cultures. Furthermore, despite this review's intent, not all the studies were set in rural and remote areas, and therefore, its findings cannot be directly adopted in rural and remote contexts, with the exception of those of the third research question. The search included a comprehensive variation of health- and technology-related search terms (Textbox 1). The search terms did not include common abbreviations, such as *eHealth* and *mHealth.* However, these absences did not likely have a significant effect on the results. In addition, health technologies develop rapidly [35]; therefore, this review's start date of 2010 may seem distant. However, a 10-year period is commonly used in qualitative reviews of eHealth technology [14,29]. As noted by Vuojärvi [79], "although technologies change rapidly, people and the ways technologies are used in everyday lives, and particularly in the learning process, do not necessarily do so."

Conclusions

This literature review provides information and practical implications for designers, health care providers, and policy makers. eHealth technology targeted at older adults should be easy to use, and adequate support and training should be provided to users [25,80]. There are plenty of barriers that prevent older adults from ably and independently using eHealth technologies at home. Special attention should be paid to the most common barriers to learning how to use and using eHealth technologies: health-related difficulties, lack of internet connectivity, and other technical problems. We would like to

emphasize the importance of considering older adults' social and cultural practices when designing and implementing eHealth technologies. Social network support and technology integration into everyday life in rural areas contribute to a successful domestication process.

Acknowledgments

This study was part of a larger research project called HARVEST—eHealth and Aging in Rural Areas: Transforming Everyday Life, Digital Competences, and Technology (2018-2021). The project is nationally supported by the Academy of Finland (318835) as part of the international JPI MYBL (Joint Programming Initiative More Years, Better Lives) call. JPI MYBL is supported by the J-AgeII. J-AgeII is funded by Horizon 2020, the EU Framework Program for Research and Innovation, under grant agreement 643850, the JPI MYBL.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), a 27-item checklist.
[PDF File (Adobe PDF File), 103 KB - jmir_v23i12e23804_app1.pdf]

Multimedia Appendix 2
An example search string for Scopus (Elsevier).
[PDF File (Adobe PDF File), 50 KB - jmir_v23i12e23804_app2.pdf]

Multimedia Appendix 3
Summary of included articles.
[PDF File (Adobe PDF File), 147 KB - jmir_v23i12e23804_app3.pdf]

References

1. 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]
2. Burholt V, Dobbs C. Research on rural ageing: where have we got to and where are we going in Europe? J Rural Stud 2012 Oct;28(4):432-446. [doi: 10.1016/j.jrurstud.2012.01.009]
3. Kilpeläinen A, Seppänen M. Information technology and everyday life in ageing rural villages. J Rural Stud 2014 Jan;33:1-8. [doi: 10.1016/j.jrurstud.2013.10.005]
4. Pruchno R. Technology and aging: an evolving partnership. Gerontologist 2019 Jan 09;59(1):1-5. [doi: 10.1093/geront/gny153] [Medline: 30629258]
5. Official records of World Health Organization, no 2. Summary report on proceedings minutes and final acts of the international health conference health in New York from 19 June to 22 July 1946. United Nations World Health Organization Interim Commission. URL: https://apps.who.int/iris/bitstream/handle/10665/85573/Official_record2_eng.pdf?sequence=1&isAllowed=y [accessed 2021-11-16]
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
8. World population ageing 2019. United Nations. 2019. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Report.pdf [accessed 2021-11-24]
9. World Health Organization. Global health and aging.: National Institute on Aging, National Institute of Health; 2011. URL: https://www.who.int/ageing/publications/global_health.pdf [accessed 2021-11-16]
10. Kröger T, Bagnato A. Care for older people in early twenty-first-century Europe: Dimensions and directions of change. In: Martinelli F, Anttonen A, Mätzke M, editors. Social Services Disrupted: Changes, Challenges and Policy Implications for Europe in Times of Austerity. New Horizons in Social Policy Series. Cheltenham, England: Elgar; 2017:201-218.
11. Murgic L, Hébert PC, Sovic S, Pavlekovic G. Paternalism and autonomy: views of patients and providers in a transitional (post-communist) country. BMC Med Ethics 2015 Sep 29;16(1):65 [FREE Full text] [doi: 10.1186/s12910-015-0059-z] [Medline: 26420014]

12. Vuorikari R, Punie Y, Carretero S, den Brande LV. DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model. Luxembourg: Publication Office of the European Union; 2016. URL: https://op.europa.eu/en/publication-detail/-/publication/bc52328b-294e-11e6-b616-01aa75ed71a1
13. Lipponen L. Information literacy as a situated and distributed activity. In: Lloyd A, Talja S, editors. Practising Information Literacy: Bringing Theories of Learning, Practice and Information Literacy Together. Wagga Wagga: Centre for Information Studies, Charles Sturt University; 2010:51-64.
14. Rasi P, Kilpeläinen A. The digital competences and agency of older people living in rural villages in Finnish Lapland. Seminar net Int J Media Technol Lifelong Learn 2015 Nov 7;11(2):149-160 [FREE Full text]
15. Haddon L. The contribution of domestication research to in-home computing and media consumption. Inf Soc 2006 Sep;22(4):195-203. [doi: 10.1080/01972240600791325]
16. Lie M, Sørensen KH. Making Technology Our Own? Domesticating Technology Into Everyday Life. Oslo, Norway: Scandinavian University Press; 1996.
17. Silverstone R, Haddon L. Design and the domestication of information and communication technologies: technical change and everyday life. In: Mansell R, Silverstone R, editors. Communication by Design: The Politics of Information and Communication Technologies. Oxford, England, UK: Oxford University Press; 1996:44-74.
18. Hynes D, Richardson H. What use is domestication theory to information systems research? In: Dwivedi YK, Lal B, Williams MD, Schneberger SL, Wade M, editors. Handbook of Research on Contemporary Theoretical Models in Information Systems. Hershey, PA: IGI Global; 2009:482-494.
19. Hynes D, Rommes E. Fitting the internet into our lives: internet courses for disadvantaged users. In: Berker T, Hartmann M, Punie Y, Ward KJ, editors. Domestication of Media and Technologies. Berkshire, England: Open University Press; 2006:123-144.
20. Scheerder AJ, van Deursen AJ, van Dijk JA. Internet use in the home: digital inequality from a domestication perspective. New Media Soc 2019 May 02;21(10):2099-2018. [doi: 10.1177/1461444819844299]
21. Silverstone R. Domesticating domestication: reflections on the life of a concept. In: Berker T, Hartmann M, Punie Y, Ward KJ, editors. Domestication of Media and Technologies. Berkshire, England: Open University Press; 2006:123-144.
22. Gubrium J, Holstein J. Introduction. In: Aging and Everyday Life. Oxford, England: Blackwell Publishers; 2000.
23. Haddon L. Domestication analysis, objects of study, and the centrality of technologies in everyday life. Can J Commun 2011 Aug 04;36(2):311-323 [FREE Full text] [doi: 10.22230/cjc.2011v36n2a2322]
24. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Oct;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
25. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
26. 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 05;16(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
27. Deutsch I, Erel H, Paz M, Hoffman G, Zuckerman O. Home robotic devices for older adults: opportunities and concerns. Comput Human Behav 2019 Sep;98:122-133. [doi: 10.1016/j.chb.2019.04.002]
28. Fang ML, Siden E, Korol A, Demestihas M, Sixsmith J, Sixsmith A. A scoping review exploration of the intended and unintended consequences of eHealth on older people: a health equity impact assessment. Human Technol 2018 Nov 30:297-323. [doi: 10.17011/ht/urn.201811224835]
29. Spann A, Stewart E. Barriers and facilitators of older people's mHealth usage: a qualitative review of older people's views. Human Technol 2018 Nov 30:264-296. [doi: 10.17011/ht/urn.201811224834]
30. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
31. Merkel S, Kucharski A. Participatory design in gerontechnology: a systematic literature review. Gerontologist 2019 Jan 09;59(1):e16-e25. [doi: 10.1093/geront/gny034] [Medline: 29788319]
32. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
33. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
34. Heyn P, Meeks S, Pruchno R. Methodological guidance for a quality review article. Gerontologist 2019 Mar 14;59(2):197-201. [doi: 10.1093/geront/gny123] [Medline: 30698698]
35. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]

36. Lee SB, Oh JH, Park JH, Choi SP, Wee JH. Differences in youngest-old, middle-old, and oldest-old patients who visit the emergency department. Clin Exp Emerg Med 2018 Dec;5(4):249-255 [FREE Full text] [doi: 10.15441/ceem.17.261] [Medline: 30571903]
37. Boquete L, Rodríguez-Ascariz JM, Amo-Usanos C, Martínez-Arribas A, Amo-Usanos J, Otón S. User-friendly cognitive training for the elderly: a technical report. Telemed J E Health 2011;17(6):456-460. [doi: 10.1089/tmj.2010.0149] [Medline: 21612520]
38. Colón-Semenza C, Latham NK, Quintiliani LM, Ellis TD. Peer coaching through mHealth targeting physical activity in people with parkinson disease: feasibility study. JMIR Mhealth Uhealth 2018 Feb 15;6(2):e42 [FREE Full text] [doi: 10.2196/mhealth.8074] [Medline: 29449201]
39. Read Paul L, Salmon C, Sinnarajah A, Spice R. Web-based videoconferencing for rural palliative care consultation with elderly patients at home. Support Care Cancer 2019 Sep;27(9):3321-3330. [doi: 10.1007/s00520-018-4580-8] [Medline: 30613908]
40. Philip L, Roberts A, Currie M, Mort A. technology for older adults: maximising personal and social interaction: exploring opportunities for eHealth to support the older rural population with chronic pain. Scott Geogr J 2015 Jan 15;131(3-4):181-193. [doi: 10.1080/14702541.2014.978806]
41. Vaskinn A, Wilsgård I, Holm A, Wootton R, Elvevåg B. A feasibility study of a telephone-based screening service for mild cognitive impairment and its uptake by elderly people. J Telemed Telecare 2013 Jan;19(1):5-10. [doi: 10.1177/1357633X12473904] [Medline: 23390214]
42. Hicken BL, Daniel C, Luptak M, Grant M, Kilian S, Rupper RW. Supporting Caregivers of Rural Veterans Electronically (SCORE). J Rural Health 2017 Jun;33(3):305-313. [doi: 10.1111/jrh.12195] [Medline: 27437642]
43. Pavic M, Klaas V, Theile G, Kraft J, Tröster G, Blum D, et al. Mobile health technologies for continuous monitoring of cancer patients in palliative care aiming to predict health status deterioration: a feasibility study. J Palliat Med 2020 May;23(5):678-685. [doi: 10.1089/jpm.2019.0342] [Medline: 31873052]
44. Blusi M, Asplund K, Jong M. Older family carers in rural areas: experiences from using caregiver support services based on Information and Communication Technology (ICT). Eur J Ageing 2013 Sep;10(3):191-199 [FREE Full text] [doi: 10.1007/s10433-013-0260-1] [Medline: 28804294]
45. Sorocco KH, Bratkovich KL, Wingo R, Qureshi SM, Mason PJ. Integrating care coordination home telehealth and home based primary care in rural Oklahoma: a pilot study. Psychol Serv 2013 Aug;10(3):350-352. [doi: 10.1037/a0032785] [Medline: 23937085]
46. Pigini L, Bovi G, Panzarino C, Gower V, Ferratini M, Andreoni G, et al. Pilot test of a new personal health system integrating environmental and wearable sensors for telemonitoring and care of elderly people at home (SMARTA project). Gerontology 2017;63(3):281-286. [doi: 10.1159/000455168] [Medline: 28099965]
47. Williams K, Pennathur P, Bossen A, Gloeckner A. Adapting telemonitoring technology use for older adults: a pilot study. Res Gerontol Nurs 2016;9(1):17-23 [FREE Full text] [doi: 10.3928/19404921-20150522-01] [Medline: 26020575]
48. Barrera TL, Cummings JP, Armento M, Cully JA, Bush Amspoker A, Wilson NL, et al. Telephone-delivered cognitive-behavioral therapy for older, rural veterans with depression and anxiety in home-based primary care. Clin Gerontol 2017;40(2):114-123. [doi: 10.1080/07317115.2016.1254133] [Medline: 28452676]
49. Luptak M, Dailey N, Juretic M, Rupper R, Hill R, Hicken B, et al. The Care Coordination Home Telehealth (CCHT) rural demonstration project: a symptom-based approach for serving older veterans in remote geographical settings. Rural Remote Health 2010;10(2):1375 [FREE Full text] [Medline: 20518592]
50. Blusi M, Kristiansen L, Jong M. Exploring the influence of internet-based caregiver support on experiences of isolation for older spouse caregivers in rural areas: a qualitative interview study. Int J Older People Nurs 2015 Sep;10(3):211-220. [doi: 10.1111/opn.12074] [Medline: 25425070]
51. Ottenberg AL, Swetz KM, Mueller LA, Gerhardson S, Mueller PS. "We as Human Beings Get Farther and Farther Apart": the experiences of patients with remote monitoring systems. Heart Lung 2013;42(5):313-319 [FREE Full text] [doi: 10.1016/j.hrtlng.2013.03.002] [Medline: 23582212]
52. Baharin H, Viller S, Rintel S. SonicAIR: supporting independent living with reciprocal ambient audio awareness. ACM Trans Comput Hum Interact 2015 Jul 09;22(4):1-23. [doi: 10.1145/2754165]
53. Garattini C, Wherton J, Prendergast D. Linking the lonely: an exploration of a communication technology designed to support social interaction among older adults. Univ Access Inf Soc 2011 Jun 7;11(2):211-222. [doi: 10.1007/s10209-011-0235-y]
54. Sadek I, Mohktari M. Nonintrusive remote monitoring of sleep in home-based situation. J Med Syst 2018 Mar 01;42(4):64. [doi: 10.1007/s10916-018-0917-6] [Medline: 29497864]
55. Holz EM, Botrel L, Kaufmann T, Kübler A. Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study. Arch Phys Med Rehabil 2015 Mar;96(3 Suppl):S16-S26. [doi: 10.1016/j.apmr.2014.03.035] [Medline: 25721543]
56. Naik AD, White CD, Robertson SM, Armento ME, Lawrence B, Stelljes LA, et al. Behavioral health coaching for rural-living older adults with diabetes and depression: an open pilot of the HOPE study. BMC Geriatr 2012 Jul 24;12:37 [FREE Full text] [doi: 10.1186/1471-2318-12-37] [Medline: 22828177]

57. Far I, Ferron M, Iberra F, Baez M, Tranquillini S, Casati F, et al. The interplay of physical and social wellbeing in older adults: investigating the relationship between physical and social interaction with virtual social environments. PeerJ Comput Sci 2015 Nov 25;1:e30. [doi: 10.7717/peerj-cs.30]
58. 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]
59. Costa CR, Iglesias MJ, Rifón LE, Carballa MG, Rodríguez SV. The acceptability of TV-based game platforms as an instrument to support the cognitive evaluation of senior adults at home. PeerJ 2017;5:e2845 [FREE Full text] [doi: 10.7717/peerj.2845] [Medline: 28070464]
60. de Garibay VG, Fernández MA, de la Torre-Díez I, López-Coronado M. Utility of a mHealth app for self-management and education of cardiac diseases in Spanish urban and rural areas. J Med Syst 2016 Aug;40(8):186. [doi: 10.1007/s10916-016-0531-4] [Medline: 27329050]
61. Parker B, Rajapakshe R, Moldovan A, Araujo C, Crook J. An internet-based means of monitoring quality of life in post-prostate radiation treatment: a prospective cohort study. JMIR Res Protoc 2015 Sep 28;4(3):e115 [FREE Full text] [doi: 10.2196/resprot.3974] [Medline: 26416584]
62. Strain T, Wijndaele K, Brage S. Physical activity surveillance through smartphone apps and wearable trackers: examining the UK potential for nationally representative sampling. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11898 [FREE Full text] [doi: 10.2196/11898] [Medline: 30694198]
63. Evangelista LS, Lee J, Moore AA, Motie M, Ghasemzadeh H, Sarrafzadeh M, et al. Examining the effects of remote monitoring systems on activation, self-care, and quality of life in older patients with chronic heart failure. J Cardiovasc Nurs 2015;30(1):51-57 [FREE Full text] [doi: 10.1097/JCN.0000000000000110] [Medline: 24365871]
64. West SP, Lagua C, Trief PM, Izquierdo R, Weinstock RS. Goal setting using telemedicine in rural underserved older adults with diabetes: experiences from the informatics for diabetes education and telemedicine project. Telemed J E Health 2010 May;16(4):405-416. [doi: 10.1089/tmj.2009.0136] [Medline: 20507198]
65. Bock C, Demiris G, Choi Y, Le T, Thompson HJ, Samuel A, et al. Engaging older adults in the visualization of sensor data facilitated by an open platform for connected devices. Technol Health Care 2016 Mar 11;24(4):541-550. [doi: 10.3233/THC-161150] [Medline: 27002473]
66. Wakefield B, Drwal K, Scherubel M, Klobucar T, Johnson S, Kaboli P. Feasibility and effectiveness of remote, telephone-based delivery of cardiac rehabilitation. Telemed J E Health 2014 Jan;20(1):32-38 [FREE Full text] [doi: 10.1089/tmj.2013.0079] [Medline: 24161003]
67. Saleh S, Farah A, El Arnaout N, Dimassi H, El Morr C, Muntaner C, et al. mHealth use for non-communicable diseases care in primary health: patients' perspective from rural settings and refugee camps. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii52-ii63 [FREE Full text] [doi: 10.1093/pubmed/fdy172] [Medline: 30307516]
68. Kamimura T, Ishiwata R, Inoue T. Medication reminder device for the elderly patients with mild cognitive impairment. Am J Alzheimers Dis Other Demen 2012 Jun;27(4):238-242 [FREE Full text] [doi: 10.1177/1533317512450066] [Medline: 22739031]
69. Mitzner TL, Savla J, Boot WR, Sharit J, Charness N, Czaja SJ, et al. Technology adoption by older adults: findings from the PRISM trial. Gerontologist 2019 Jan 09;59(1):34-44 [FREE Full text] [doi: 10.1093/geront/gny113] [Medline: 30265294]
70. Olsson T, Viscovi D. Warm experts for elderly users: who are they and what do they do? Human Technol 2018 Nov 30;14(3):324-342. [doi: 10.17011/ht/urn.201811224836]
71. Airola E, Rasi P. Domestication of a robotic medication-dispensing service among older people in Finnish Lapland. Human Technology 2020;16(2):117-138. [doi: 10.17011/ht/urn.202008245639]
72. Hasan H, Linger H. Enhancing the wellbeing of the elderly: social use of digital technologies in aged care. Educ Gerontol 2016 Jun 28;42(11):749-757. [doi: 10.1080/03601277.2016.1205425]
73. Tsai HY, Shillair R, Cotten SR, Winstead V, Yost E. Getting grandma online: are tablets the answer for increasing digital inclusion for older adults in the U.S.? Educ Gerontol 2015;41(10):695-709 [FREE Full text] [doi: 10.1080/03601277.2015.1048165] [Medline: 26877583]
74. Brenes GA, Danhauer SC, Lyles MF, Hogan PE, Miller ME. Telephone-delivered cognitive behavioral therapy and telephone-delivered nondirective supportive therapy for rural older adults with generalized anxiety disorder: a randomized clinical trial. JAMA Psychiatry 2015 Oct;72(10):1012-1020 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1154] [Medline: 26244854]
75. Do HM, Pham M, Sheng W, Yang D, Liu M. RiSH: a robot-integrated smart home for elderly care. Rob Auton Syst 2018 Mar;101:74-92. [doi: 10.1016/j.robot.2017.12.008]
76. Son H, Kim H. A pilot study to test the feasibility of a home mobility monitoring system in community-dwelling older adults. Int J Environ Res Public Health 2019 Apr 29;16(9):1512 [FREE Full text] [doi: 10.3390/ijerph16091512] [Medline: 31035678]
77. Fossum M, Alexander GL, Ehnfors M, Ehrenberg A. Effects of a computerized decision support system on pressure ulcers and malnutrition in nursing homes for the elderly. Int J Med Inform 2011 Sep;80(9):607-617. [doi: 10.1016/j.ijmedinf.2011.06.009] [Medline: 21783409]

78. Haddon L. Roger Silverstone's legacies: domestication. New Media Soc 2016 Jun 30;9(1):25-32. [doi: 10.1177/1461444807075201]
79. Vuojärvi H. Conceptualising Personal and Mobile Learning Environments in Higher Education: Focus on Students' Perspective. Academic Dissertation. Acta Universitatis Lapponiensis 266. Rovaniemi, Finland: Lapland University Press; 2013.
80. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. 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]

Abbreviations

ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka, G Eysenbach; submitted 24.08.20; peer-reviewed by C Cheng, D Estrin; comments to author 25.10.20; revised version received 23.11.20; accepted 14.10.21; published 02.12.21.

Please cite as:
Airola E
Learning and Use of eHealth Among Older Adults Living at Home in Rural and Nonrural Settings: Systematic Review
J Med Internet Res 2021;23(12):e23804
URL: https://www.jmir.org/2021/12/e23804
doi:10.2196/23804
PMID:34860664

Review

Establishing a Working Definition of User Experience for eHealth Interventions of Self-reported User Experience Measures With eHealth Researchers and Adolescents: Scoping Review

Amanda S Newton[1], PhD; Sonja March[2], PhD; Nicole D Gehring[1], MSc; Arlen K Rowe[2], PhD; Ashley D Radomski[3,4], PhD

[1]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

[2]School of Psychology and Counselling, Centre for Health Research, University of Southern Queensland, Springfield Central, Australia

[3]Knowledge Institute for Child and Youth Mental Health and Addictions, Ottawa, ON, Canada

[4]CHEO (Children's Hospital of Eastern Ontario) Research Institute, Ottawa, ON, Canada

Corresponding Author:
Amanda S Newton, PhD
Department of Pediatrics
University of Alberta
3-526 Edmonton Clinic Health Academy
11405-87 Avenue
Edmonton, AB, T6G 1C9
Canada
Phone: 1 7802485581
Email: mandi.newton@ualberta.ca

Abstract

Background: Across eHealth intervention studies involving children, adolescents, and their parents, researchers have measured user experience to assist with intervention development, refinement, and evaluation. To date, no widely accepted definitions or measures of *user experience* exist to support a standardized approach for evaluation and comparison within or across interventions.

Objective: We conduct a scoping review with subsequent Delphi consultation to identify how user experience is defined and measured in eHealth research studies, characterize the measurement tools used, and establish working definitions for domains of user experience that could be used in future eHealth evaluations.

Methods: We systematically searched electronic databases for published and gray literature available from January 1, 2005, to April 11, 2019. We included studies assessing an eHealth intervention that targeted any health condition and was designed for use by children, adolescents, and their parents. eHealth interventions needed to be web-, computer-, or mobile-based, mediated by the internet with some degree of interactivity. We required studies to report the measurement of *user experience* as first-person experiences, involving cognitive and behavioral factors reported by intervention users. We appraised the quality of user experience measures in included studies using published criteria: *well-established*, *approaching well-established*, *promising*, or *not yet established*. We conducted a descriptive analysis of how user experience was defined and measured in each study. Review findings subsequently informed the survey questions used in the Delphi consultations with eHealth researchers and adolescent users for how user experience should be defined and measured.

Results: Of the 8634 articles screened for eligibility, 129 articles and 1 erratum were included in the review. A total of 30 eHealth researchers and 27 adolescents participated in the Delphi consultations. On the basis of the literature and consultations, we proposed working definitions for 6 main user experience domains: acceptability, satisfaction, credibility, usability, user-reported adherence, and perceived impact. Although most studies incorporated a study-specific measure, we identified 10 well-established measures to quantify 5 of the 6 domains of user experience (all except for self-reported adherence). Our adolescent and researcher participants ranked perceived impact as one of the most important domains of user experience and usability as one of the least important domains. Rankings between adolescents and researchers diverged for other domains.

Conclusions: Findings highlight the various ways in which user experience has been defined and measured across studies and what aspects are most valued by researchers and adolescent users. We propose incorporating the working definitions and available measures of user experience to support consistent evaluation and reporting of outcomes across studies. Future studies can refine

the definitions and measurement of user experience, explore how user experience relates to other eHealth outcomes, and inform the design and use of human-centered eHealth interventions.

(J Med Internet Res 2021;23(12):e25012) doi:10.2196/25012

KEYWORDS

eHealth; internet; design; development; user experience; health care; scoping review; Delphi

Introduction

Background

Over the past 15 years, the number of eHealth interventions available for use by children, adolescents, and their parents has grown considerably. A commonly used approach to eHealth intervention development involves human-centered design (also known as patient- or user-centered design) [1,2]. This approach includes the active participation of intervention users—children, adolescents, and parents—in the intervention design and development process. By including user perspectives and input into intervention design, the likelihood that an intervention will be easy to use and be compatible with the user and their individual context, and therefore deemed useful, is improved [3-5]. More recently, the importance of users' involvement in intervention evaluation has been recognized, with measures of *user experience* included in evaluations to identify whether and how an eHealth intervention meets the preferences and needs of the users. Understanding user experience can more or less reflect the quality of human-centered design principles associated with an intervention.

The term *user experience* initially arose in the field of human-computer interaction and technology design and was broadly defined as, "a person's perception and responses that result from the use or anticipated use of a product, system or service" [6]. To date, across eHealth studies, a wide range of definitions and concepts have been used to evaluate user experience, such as satisfaction, acceptability, adherence, engagement, and usability, with an eHealth intervention [7-14]. Similarly, user experience data collection methods have also varied, such as with the use of self-report questionnaires, in-person or telephone-based interview guides, or different types of automatic data capture of users' interactions with an intervention [15]. These variations suggest that user experience may be a multidimensional concept with several important constructs to define and measure within an eHealth intervention, and a consensus among researchers is yet to be reached. Similar to how the need to define, standardize, and measure adherence has been mounting in recent years [16,17], a need to converge on a common understanding of user experience is also becoming more apparent. A set of accepted domains, definitions, and evaluation measures used in eHealth intervention development and evaluation would benefit children, adolescents, and parents by allowing them to compare user experiences between multiple interventions and inform decisions about their own eHealth intervention use. These accepted approaches would also provide guidance to researchers in the eHealth field and allow for continued advancement and improvement of the study of user experience and other eHealth outcomes (eg, intervention effectiveness, user safety, and user empowerment), intervention design components (eg, content and technological features), and factors that can influence the intervention experience of users (eg, context of use and user expectations).

This Study

This study involves two phases: a scoping review and Delphi consultations. Our decision to conduct a review plus consultation reflects a hermeneutic position that user experience cannot be fully understood without examining it in its current context (existing literature) and the meanings attributed to it (Delphi consultations). The scoping review includes diverse literature to identify how user experience has been defined and measured in eHealth research studies of children, adolescents, and parents. These findings subsequently informed the development of surveys used in the Delphi consultations with eHealth researchers and adolescent users of an eHealth intervention, which focused on establishing a working definition of user experience (and the domains that it may encompass) and developing recommendations for measuring user experience in future evaluations of eHealth interventions.

Methods

Study Design

We followed the scoping review framework proposed by Arksey and O'Malley [18] with a Delphi consultation recommended by Levac et al [19]. Reporting of the review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) Extension for Scoping Reviews checklist [20]. The consultation exercise followed synthesis of findings from the literature. Approvals from the Health Research Ethics Board at the University of Alberta and Human Research Ethics Committee at the University of Southern Queensland were received for the Delphi consultation. The approved study protocol is available upon request.

Development of the Search Strategy

The search strategy (Multimedia Appendix 1) was developed using an iterative process. First, we developed a list of search terms using key concepts and terms from a convenience sample of indexed studies published in various years that examined user experience with an eHealth intervention. A research librarian provided input on appropriate filters, such as Medical Subject Headings terms, and modified these terms to comply with different databases. For this review, we were interested in identifying both published and unpublished English language studies of user experiences, and we sought to include studies made available between January 1, 2005, and April 11, 2019. We included the literature published since 2005 to focus on contemporary studies of eHealth technologies (eg, mobile apps, desktop-based, and multisession or single-session interventions).

The search terms and parameters were tested for sensitivity, determined by whether the search strategy successfully filtered the 63 citations that were manually selected a priori (see Multimedia Appendix 2 for the list of test citations). We then conducted 2 rounds of preliminary screening to further refine the strategy. The finalized search strategy was peer-reviewed before implementation.

Search Strategy

We identified studies from the following databases: Ovid MEDLINE, PsycINFO, CINAHL, EBM Reviews (Cochrane Database of Systematic Reviews, ACP Journal Club, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment, and NHS Economic Evaluation Database), Cochrane Central Register of controlled trials, and ClinicalTrials.gov. We searched Google Scholar from January 2005 to April 2019 and conference proceedings of the International Society for Research on Internet Interventions from January 2016 to April 2019, as there are no archives of International Society for Research on Internet Interventions previous to 2016. We reviewed the reference lists from reviews (systematic, narrative, etc) to identify additional, potentially relevant studies.

Criteria for Considering Studies for the Scoping Review

We included studies of any design that assessed user experience with an eHealth treatment or prevention intervention designed for children or adolescents (aged ≤19 years). Studies with a sample that contained young adults could be included in the review if the mean age of participants was reported to be ≤19 years. eHealth interventions could target any health condition but needed to be web-, computer-, or mobile-based, mediated by the internet and include some degree of interactivity. Studies of telehealth interventions were not included. Studies could assess the eHealth user experience of children, adolescents, or parents. Given the wide range of pre-existing definitions and measurement approaches used to evaluate eHealth intervention user experience, multiple domains could be included in the user experience evaluation, such as cognitive factors (ie, beliefs, attitudes, and intention; such as satisfaction and acceptability of the intervention) or behavioral factors (ie, how the intervention was used, such as self-reported adherence to and engagement with the intervention) related to the use of an eHealth intervention. To be included, studies needed to report the measurement of *user experience* as first-person experiences reported by eHealth intervention users (parents, children, and adolescents). Studies that only reported indirect user data (ie, proxy report by a parent whose child used a program and intervention metadata [number of sessions completed]), which do not reflect the user's subjective experience, were excluded. Studies also needed to detail the evaluation measures (eg, tool, instrument, or interview questions) used to collect user experience data so that we could identify how user experience was defined and measured. Studies that did not detail evaluation questions but referenced an original publication of the evaluation measure were included if we could obtain the referenced publication to extract information.

Screening for Article Eligibility

We organized and screened identified studies using EndNote (Clarivate Analytics) bibliographic management software. In pairs, 3 reviewers (authors NDG and AKR and review contributor Marcus O'Neill) independently screened the title and abstract of articles, classifying each as *relevant*, *irrelevant*, or *unclear* using the predetermined inclusion and exclusion criteria. To assess the clarity of the criteria for each reviewer during screening, we calculated the interrater agreement for screening outcomes for the first 100 articles using the κ statistic [21]. The agreement was *substantial* (κ=0.61). We wanted interrater agreement to be ≥0.80, indicating an *almost perfect* agreement [22], so reviewers met to review the screening criteria alongside the articles for which they disagreed and sought consensus on the screening outcome for each article. Agreement increased to *almost perfect* (κ=0.81) for the next set of 100 articles and therefore the screening progressed. At this point, reviewers divided the remaining articles to be screened. Articles screened as relevant or unclear underwent independent screening and discussion by each reviewer pair to determine study inclusion or exclusion. The reviewers contacted the primary authors of 9 articles when additional information was needed to determine eligibility. The reviewers documented the articles that were excluded after full-text review to ensure transparency and replicability.

Data Extraction

Process

Data were extracted into a standardized spreadsheet. The spreadsheet underwent pilot testing with 3 independent reviewers (authors NDG and AKR and review contributor MO) who extracted data from the first 5 included studies to ensure that the spreadsheet was adequate in scope and that consensus was achieved on data categorization. Subsequently, each reviewer extracted data from one-third of the remaining included studies. Each reviewer verified the accuracy and completeness of the other reviewers' respective thirds. Data extraction discrepancies were resolved through consensus and third-party consultation (author ASN). Corresponding authors were contacted when reporting was unclear or details were lacking in the article.

Data Extracted for Analysis

We extracted the following information from the studies:

1. General information on participants (age [range] and intervention [user or respondent]) and the eHealth intervention (name, mode of delivery, target population or health condition and duration or frequency).

2. How first-person user experience was defined in studies, which included looking for definitions and terms of user-reported experiences as well as extracting individual questions used to measure user experience. We then compared author-reported user experience definitions to a priori definitions and the identified measures were categorized into 6 domains: satisfaction, acceptability, credibility, impact, adherence, and use. The domains were based on a preliminary literature review of user experiences in eHealth studies that was conducted by

one of the authors [23]. All tools fit into one or more of the 6 domains. The original working definitions for the 6 domains are presented in Multimedia Appendix 3.

3. Major characteristics of the evaluation measures used to assess user experience, including its purpose and scope, delivery time points, type of respondent (child, adolescent, or parent), administration approach (web-, telephone-, or paper-based or face-to-face interview), the number of items and item-response format (eg, Likert scale or open-ended questions), and any notations by study authors regarding limitations of the evaluation measures and recommendations for future measurement or evaluation.

4. Information on the measure's psychometrics was extracted, if available, including information on measure validity (face, content, construct, or criterion), reliability (internal consistency, interrater, or test-retest), and findings from a factor analysis. In studies where an evaluation measure was referenced, the original reference was reviewed and psychometric data were extracted, if available.

Quality Assessment

Two independent reviewers (NDG and AKR) assessed the quality of the evaluation measures reported in the studies and met to resolve discrepancies through consensus. The reviewers used 3 criteria developed by the Society for Pediatric Psychology Assessment Task Force [24]. The first criterion was the availability of details on the instrument or measure to allow evaluation and replication. This involved the reviewers confirming whether a measure was available for review in published or gray literature; study authors could have also included their measure as supplementary material to their publication. The second criterion concerned the availability of reliability and validity data for the instrument or measure. This could include psychometric data (eg, for surveys or rating scales) and data from pilot testing (eg, face and content validity or interview guide reliability for author-developed interviews). The third criterion, use of the instrument or measure by multiple, independent investigative teams as described in peer-reviewed articles, necessitated the measure or tool (including a qualitative interview guide) to have been used by more than one group.

Using the abovementioned criteria, measures were classified into 4 categories: *well-established*, *approaching well-established*, *promising*, or *not yet established*. We rated a measure as *well-established* if we could identify 2 peer-reviewed articles with very good detail of the measures and good to strong or excellent published information on both validity and reliability. We rated a measure as *approaching well-established* if we could identify 2 peer-reviewed articles with very good detail of the measure and with published information on validity *or* reliability either missing or presented in vague or poor to moderate values. We rated a measure as *promising* if we could identify 1 peer-reviewed article with sufficient detail of the measure (eg, some, but not all of the questions present) and with published information on validity *or* reliability either missing or presented in vague or poor to moderate values. Although not included in the original task force rating scheme, we rated a measure as *not yet established* if we could identify 1 peer-reviewed article with sufficient detail of the measure but with published validity or reliability information not available. The quality of validity and reliability data were interpreted using a guide presented by Phillips et al [25] (Multimedia Appendix 4).

Data Analysis

Evidence tables and a bar graph were developed to aggregate findings into descriptive and thematic summaries [5]. Descriptive summaries include information on study, participant, and eHealth intervention characteristics, user experience measures, and related psychometric statistics. Thematic summaries include grouping measures according to the quality assessment categories used to define the measures: *well-established*, *approaching well-established*, *promising*, or *not yet established*.

Delphi Consultation Process

The Delphi consultation phase of the scoping review was a stepwise process involving multiple, structured rounds of surveys to gain consensus [26-28] on how user experience should be defined and measured.

Participants

We sought input from 2 groups of individuals: researchers of the studies included in our review who had published eHealth intervention user experience evaluations and adolescents (aged 16-18 years) currently using an eHealth intervention.

We identified and contacted researcher participants using the published email contact information of lead or corresponding authors of studies included in the scoping review. We used snowball sampling so that contacted authors could also recommend colleagues with relevant expertise who may be interested in participating [26]. Each potential participant received an email invitation to participate along with an information sheet describing the Delphi consultation; those who completed the survey were considered to have given implied consent. The Delphi consultation with researchers was held over 7 months between September 2019 and March 2020.

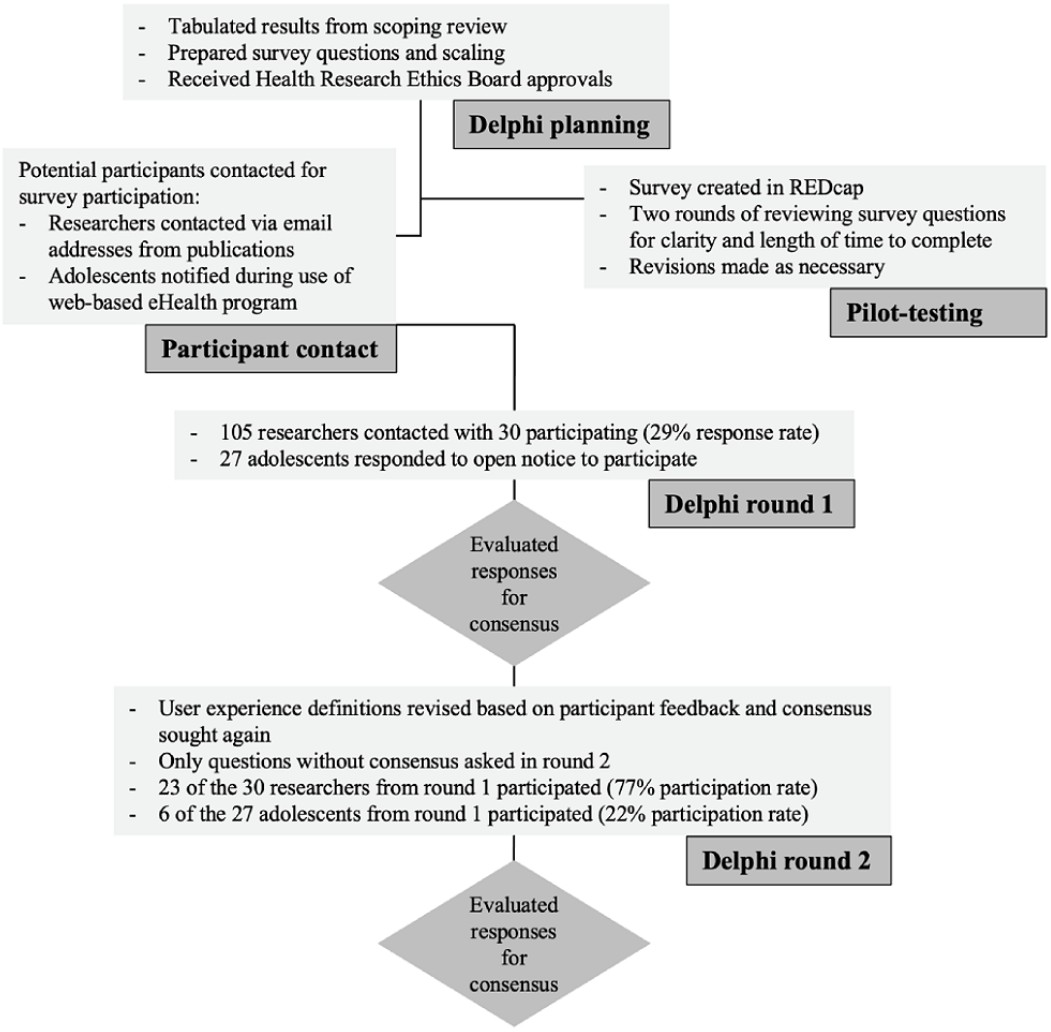
For practical and feasibility reasons, we recruited adolescent participants among current users of an evidence-based eHealth intervention, the web-based BRAVE Self-Help program [12,29], who had previously consented to be contacted for future research studies. Recruitment involved a pop-up invitation that appeared when users logged into the BRAVE Self-Help program throughout a 6-month period from June 7, 2019, to December 10, 2019. Interested participants were directed to an external site and invited to read a separate information sheet regarding the study and provide informed consent to participate in the Delphi consultation. After providing consent on the web, participants completed a survey that was used for the Delphi consultation.

Process

The process that we followed with each participant group along with the response and participation rates is presented in Figure 1. Broadly, each survey included questions that sought consensus on participants' opinions on the importance of the user experience domains and definitions used in the scoping review (Multimedia Appendix 3), additional domains for

consideration, and the appropriateness of measures used across eHealth studies of user experience. Consensus on responses to each survey question was defined as having ≥80% agreement [26].

Figure 1. Delphi consultation process. REDCap: Research Electronic Data Capture.



Survey Development and Scaling

Our team pilot-tested the surveys used in round 1 for both participant groups. We reviewed each survey for face validity, clarity, cohesiveness, flow, and completion time and made changes as needed (eg, modifying the wording of questions and changing the order of questions).

Researcher participants were first asked questions regarding their demography (age, gender, professional position, country of residence, and experience in developing and measuring user experience) followed by a series of questions on their level of agreement with the preliminary definitions of the proposed user experience domains (satisfaction, acceptability, credibility, impact, adherence, and use); initial definitions were based on a preliminary literature review that informed this scoping review as having accepted definitions or domains for measuring user experience. In round 1, an open text box for respondents to provide text supporting their answers was included. We originally wanted to survey participants for their opinions regarding the appropriateness of the user experience measures used across eHealth as well as findings from the review itself. We used a 6-point Likert scale ranging from *strongly agree* to *strongly disagree* to measure participants' level of agreement. In round 1, we also included an open text box for suggestions and comments regarding proposed definitions of identified domains and for those domains not identified in the survey. For proposed domains that could be measured at multiple time points, we also asked participants to indicate their preferred timing of measurement (before, during, or after the intervention). Researcher participants were then asked to rank the importance of the domains relative to one another using a 6-point Likert scale ranging from *most important* to *least important* when measuring user experience. Participants were also asked in both rounds to indicate the importance of studies to measure user experience; however, we realized that definitions would need to be determined before this could occur.

Adolescent participants were asked questions regarding their demography (age, gender, and frequency of using web-based health interventions) followed by questions on their level of

agreement with how important it is for researchers to ask them about each of the 6 user experience domains when using an eHealth intervention. Given the complexity and technical nature of the domains presented, adolescents were provided with lay descriptions of each of the domains and asked to rate how important they felt each domain was rather than commenting on the definition (as we did with the expert researcher sample). In this way, adolescents were able to provide input into the domains that were most important from their perspective. We used a 6-point Likert scale ranging from *extremely important* to *not at all important* to measure participants' level of agreement and also included an open text box in round 1 so that participants could identify reasons for why they felt a particular domain was important or not. Adolescent participants were then asked to rank the importance of the domains relative to one another from *most important* to *least important*. In round 1, an open text box was available for adolescents to add any other comments they had on the presented domains and any other aspects of user experience they thought were important or missing.

Consultation Rounds

We conducted independent Delphi consultations with each participant group—2 rounds with researcher participants and 2 rounds with adolescent participants—using a survey tailored to each group. The same participants participated in both rounds as the intent was to gain consensus (agreement) within the 2 participant groups. Given the nature of the questions being asked, we felt that 2 rounds were sufficient to capture participants' opinions. In round 2, we presented the summary of responses for survey questions where consensus was not achieved in round 1 or where comments and feedback for a survey question necessitated further clarification. This approach allowed each participant in round 2 to express their opinion after observing and reflecting on the opinions of other researchers or adolescents. The purpose of this approach was to decrease the degree of dispersion or increase the degree of consensus in participants' answers. The consultation exercise concluded after 2 rounds irrespective of whether consensus was reached on all survey items.

We used REDCap (Research Electronic Data Capture), a secure web-based platform [30], to administer the electronic surveys to researcher participants and the University of Southern Queensland Survey Tool based on Lime Survey and hosted on secure University of Southern Queensland servers for adolescent participants. Participants spent 15-30 minutes (researchers) or 10-15 minutes (adolescents) to complete the survey in each round. To maximize the response rate for each round, nonrespondents were sent an email reminder about the survey after every 7 days until 3 contact attempts had been made. Participants answered questions anonymously so that individual opinions did not influence other participants' opinions [26]. Adolescent participants who completed the surveys were given the opportunity to enter a draw to win one of the 10 vouchers valued at Aus $40 (US $28.92) to be drawn at completion of the study. Researchers who completed the surveys were given the opportunity to enter a draw to win a CAD $150 (US $118.14) electronic gift card.

Data Analysis

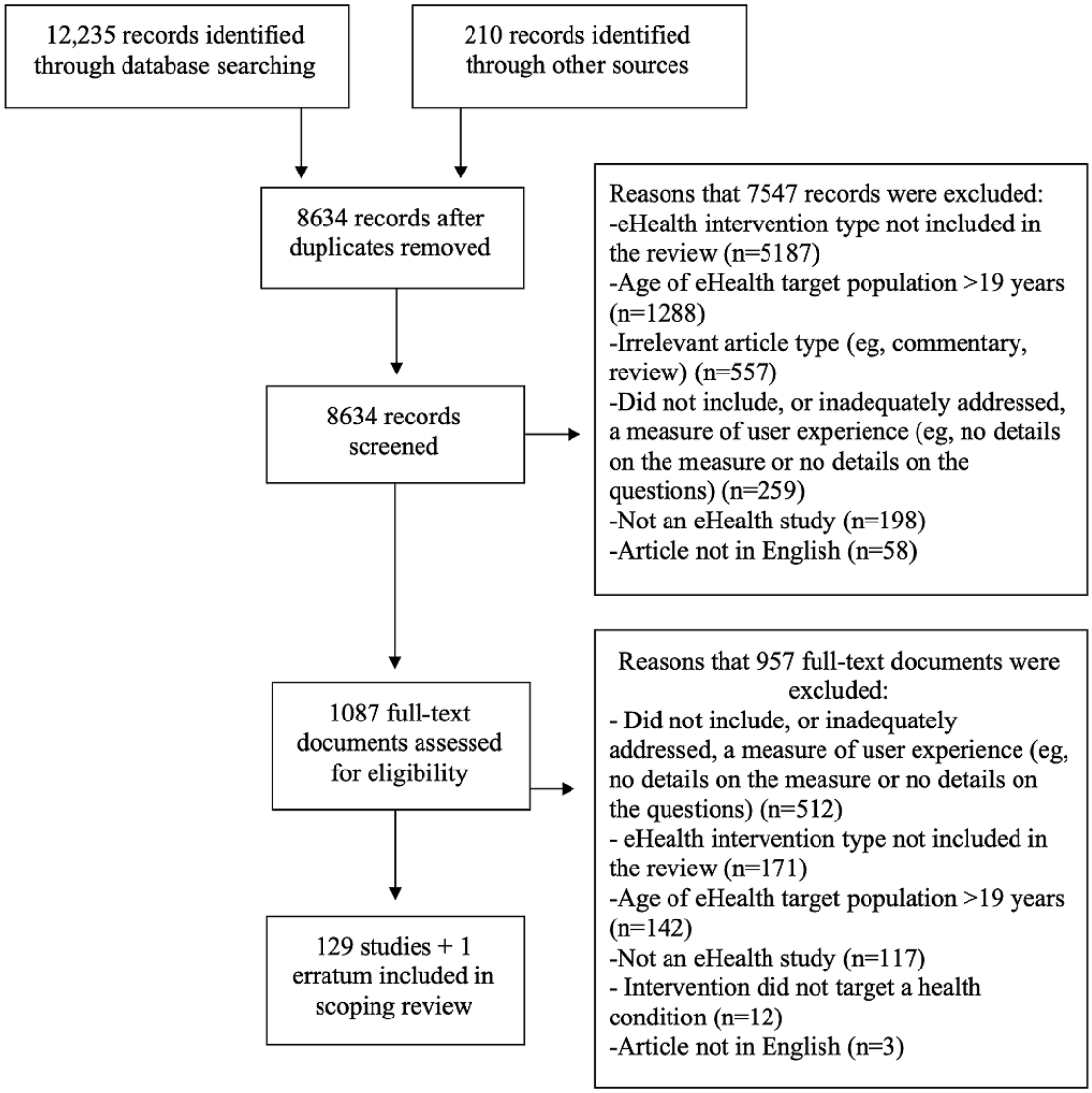
In round 1, we calculated the response rate for researcher participants only, as we were not able to determine the number of adolescent invitees from the open invitation to participate. In round 2, we calculated the participation rate for both researchers and adolescents using the denominator from round 1. For both rounds, we generated descriptive statistics (frequencies and percentages) to determine the level of agreement (consensus) among participants for each survey question. This involved grouping the responses at each end of the Likert scales (eg, grouping *strongly agree* with *agree* and *disagree* with *strongly disagree*). We also reviewed the responses to middle Likert categories (eg, *agree*/*disagree* and *slightly agree*/*slightly disagree*) to identify the range of opinions. Among researcher participants, following round 1, we collated the open-text answers or feedback on definitions and used this text to revise the definitions for the domains of user experience. We revised each domain definition using suggestions from participants irrespective of whether consensus was reached on the definition in round 1. The rationale was that the suggestions added critical details and improvements to each definition. The revised definitions were then presented for consensus in round 2 during which participants were asked to indicate their level of agreement with the revised definitions. We used IBM SPSS (version 26) for all analyses.

Results

Literature Search and Selection

The search strategy identified 8634 unique citations. Of these citations, 1087 were considered potentially relevant based on their title and abstract (Figure 2). After full-text review, 129 articles and 1 erratum met inclusion criteria and were included in the review.

Figure 2. Literature search flow diagram.



Description of Included Studies

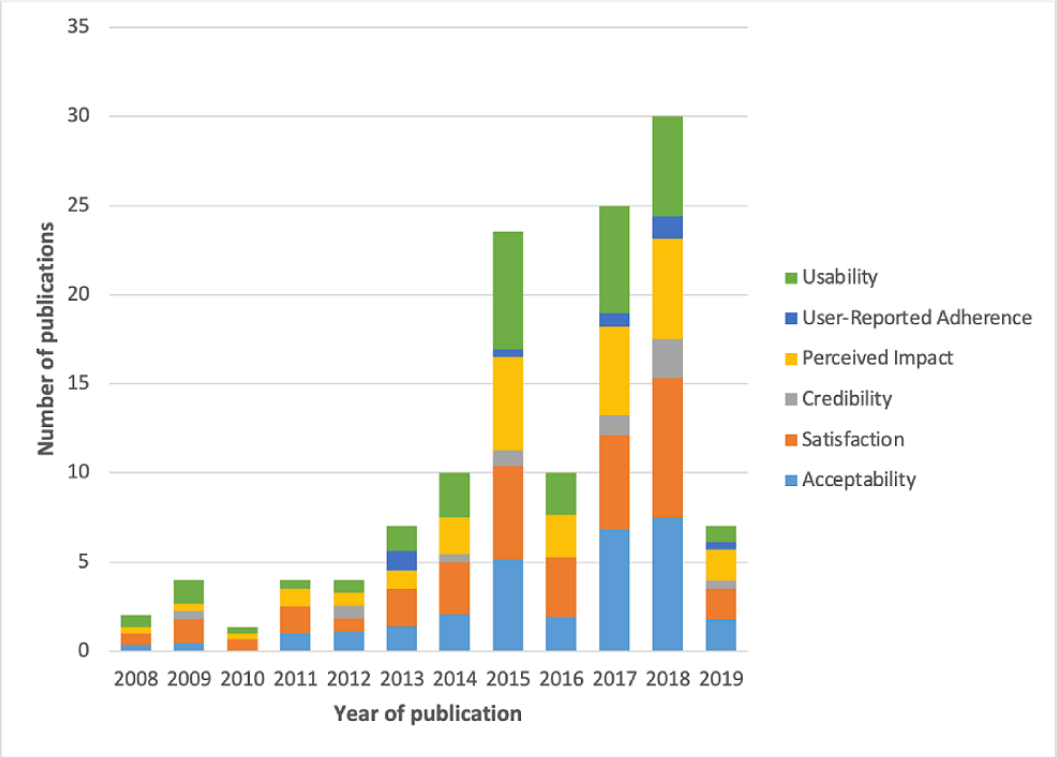
A summary of the general characteristics of 129 eHealth studies that evaluated user experience is presented in Table 1. Most studies were published in 2018 (30/129, 23.3%), 2017 (25/129, 19.4%), and 2015 (24/129, 18.6%). The most commonly measured user experience domains were acceptability, usability, and satisfaction; this trend occurred across years of publication (Figure 3). Additional details for each study are presented in Multimedia Appendix 5 [7-14,31-152], grouped by year of publication to examine trends in the domains measured over time.

Table 1. Summary of the eHealth studies that measured user experience (N=129).

Characteristics	Studies within scoping review, n (%)
eHealth user[a]	
Children aged ≤9 years	26 (20.2)
Adolescents aged 10-19 years	118 (91.5)
Young adults up to 24 years	25 (19.4)
Parents	33 (25.6)
Type of eHealth intervention	
Web-based	85 (65.9)
Mobile-based	44 (34.1)
Tablet-based	11 (8.5)
User experience domain that was measured	
Satisfaction	86 (66.7)
Acceptability	77 (59.7)
Credibility	17 (13.2)
Perceived impact	66 (51.2)
User-reported adherence	11 (8.5)
Usability	74 (57.4)

[a]Age categories defined using World Health Organization definitions [153].

Figure 3. The type and frequency of user experience domains measured across eHealth studies over time. The domains named in this figure reflect the agreed-on terminology that resulted from the Delphi consultation with researchers.



User Experience Measures

Research teams used 128 unique user experience evaluation measures in the 129 published studies included in this review. Details of the quality assessment outcomes using the Society for Pediatric Psychology Assessment Task Force [24] criteria are provided in Multimedia Appendix 6 [7-14,31-111,113-121,123-152,154,155]. Of the 128 measures,

10 (7.8%) were assessed to be *well-established* measures (Table 2) and 5 (3.9%) were assessed to be *approaching well-established* (Table 3). These measures were used in research studies to primarily capture user experiences related to satisfaction with and acceptability of eHealth interventions and program usability and perceived impact of an intervention. Evaluation measures used in research studies that were assessed to be *promising* (13/129, 10.1%) are presented in Multimedia Appendix 7 [75,90,97,99,107,109,132,136,139,152]. The remaining 100 measures identified were assessed as *not yet established* and primarily represented studies in which author-developed evaluation questions were used. Information on evaluation measures used in research studies that were assessed to be *not yet established* is available upon request to the corresponding author.

Table 2. Well-established evaluation measures of user experience.

Measure name and targeted user experience domain	Format and administration features	Psychometric properties		eHealth study
		Validity	Reliability	
SUS[a]; usability	10 items; 4- and 5-point Likert scales; administration: paper- and web-based and telephonic	Two-factor scale [156]; usable (8 items), α=.91; learnable (2 items), α=.70; overall α=.92	Internal consistency; 10 years of SUS samples: α=.91 [157]; eHealth study sample: α=.86 [31], α=.95 [32]	[14,31-39]
SUS (Portuguese version); usability	10 items; 5-point Likert scale; administration: paper-based	Construct validity with PSSUQ[b] [158]: *r*=0.70	Interrater reliability; Portuguese validation sample [158]: intraclass correlation coefficient=0.36 with modest agreement between ratings (76.67%)	[40]
Client Satisfaction Questionnaire 8; acceptability, satisfaction, and usability	8 items; 4-point Likert scale; administration: paper- and web-based	Criterion-related validity; other measures of satisfaction [159]: *r*=0.60-0.80	Internal consistency across 9 studies [159]: α=.83-.93; eHealth study sample: α=.92 [41]	[41-44]
CEQ[c]; credibility and perceived impact satisfaction	6 items; 9-point Likert scale and 0%-100% scale; administration: not reported	Two-factor scale [160]: expectancy (3 items), eigenvalue=3.42; credibility (3 items), eigenvalue=1.53; 2 factors accounted for 82.46% of the total variance	Internal consistency; CEQ validation across 3 studies [160]: expectancy, α=.79-.90; credibility, α=.81-.86; overall, α=.84-.85; test-retest reliability, CEQ validation across 3 studies [160]: expectancy, α=.82; credibility, α=.75	[41]
GEQ[d]; acceptability and satisfaction	33 items; 5-point Likert scale; administration: web-based	Five-factor scale; GEQ validation study [161]: factor 1 (5 items); factor 2 (7 items); factor 3 (4 items); factor 4 (5 items); factor 5 (4 items); all correlation coefficients >0.30	Internal consistency; GEQ development sample [162]: α=.81	[7,10]
PSSUQ also known as the Computer Systems Usability Questionnaire[e]; acceptability, perceived impact, satisfaction, and usability	19 items; 7-point Likert scale; administration: web-based	Three-factor scale; correlation coefficients on 5 years of PSSUQ samples [154]: system usefulness and informational quality, *r*=0.70; system usefulness and interface quality, *r*=0.70; informational quality and interface quality, *r*=0.60; factors scores shared 36% to 50% of the variance	Internal consistency; 5 years of PSSUQ samples [154]: system usefulness, α=.96; informational quality, α=.92; interface quality, α=.83; overall, α=.96	[45,46]
SSS[f]; satisfaction	5 items; 4-point Likert scale and open ended; administration: telephonic	One-factor scale factor loadings[g] [163]: Youth version, α=.77-.90; Parent version, α=.71-.83	Internal consistency; SSS development samples [163]: youth α=.86; parent α=.85; eHealth study sample across different time points [47]: α=.77-.95	[47]
TEI-SF[h]; acceptability and perceived impact satisfaction	9 items; 5-point Likert scale; administration: paper-based	2 factor scale [164]: acceptability (8 items), α=.49-.93, 57% of total item variance; discomfort (1 item), α=.82, 12% of total item variance	Internal consistency; TEI-SF development samples: α=.94 [165], α=.85 [164]	[48]
USE[i] questionnaire; acceptability, satisfaction, and usability	19 items (4 subscales); 7-point Likert scale; administration: web-based	Criterion-related validity [166]; compared with SUS (2 evaluations); usefulness: r_1=0.60, r_2=0.69; ease of learning: r_1=0.71, r_2=0.78; ease of use: r_1=0.78, r_2=0.81; satisfaction: r_1=0.66, r_2=0.71	Internal consistency; USE development sample [166]: α=.98; eHealth study sample [32]: usefulness, α=.90; ease of learning, α=.98; ease of use, α=.95; satisfaction, α=.96	[32]
WAI-SR[j]; credibility and perceived impact	12 items; 5-point Likert scale; administration: web-based	Three-factor scale; correlation with WAI-SR within 2 samples (S1 and S2) [167]: goal (4 items), α=.89 (S1), α=.87 (S2); task (4 items), α=.87 (S1), α=.90 (S2); bond (4 items), α=.86 (S1), α=.84 (S2); overall, α=.95 (S1), α=.94 (S2)	Internal consistency; WAI-SR development within 2 samples [167]: goal, α=.87 (S1), α=.85 (S2); task, α=.85 (S1), α=.87 (S2); bond, α=.90 (S1), α=.85 (S2); overall, α=.91 (S1), α=.92 (S2); eHealth study sample [41]: α=.95	[41,49]

[a]SUS: System Usability Scale.

[b]PSSUQ: Poststudy System Usability Questionnaire.

[c]CEQ: Credibility Expectancy Questionnaire.

[d]GEQ: Game Experience Questionnaire.

[e]The Computer Systems Usability Questionnaire and PSSUQ are the same questionnaire; the only difference is that the Computer Systems Usability Questionnaire wording is appropriate for use in field settings or surveys rather than in a scenario-based usability evaluation [154,155].

[f]SSS: Satisfaction with Services Scale.

[g]Factor loading: correlation coefficient for the variable and factor.

[h]TEI-SF: Treatment Evaluation Inventory-Short Form.

[i]USE: Usefulness, Satisfaction, and Ease of use.

[j]WAI-SR: Working Alliance Inventory: Revised Short form.

Table 3. Evaluation measures assessed to be approaching well-established.

Measure name and targeted user experience domain	Format and administration features	Psychometric properties		eHealth study
		Validity	Reliability	
Client Satisfaction Scale; perceived impact and satisfaction	10 items; 5-point Likert scale; administration: not reported	Not reported	Internal consistency; eHealth study sample [50]: Child scale, α=.75; Parent scale, α=.85	[50-52]
Standardized SUMI[a]; acceptability, satisfaction, and usability	55 items; 3-point Likert scale and open ended; administration: paper-based	Not reported	Internal consistency; SUMI development sample [168]: global subscale, α=.92; efficiency subscale, α=.81; affect subscale, α=.85; helpfulness subscale, α=.83; control subscale, α=.71; learnability subscale, α=.82	[53]
WAMMI[b]; acceptability and usability	20 items; 5-point Likert scale; administration: not reported	Not reported	Internal consistency; WAMMI development sample [169]: α=.96	[54]
Author-adapted TEI-SF[c]; acceptability and satisfaction	11 items; 5-point Likert scale; administration: paper-based	Not reported for author adaption of TEI-SF	Internal consistency; eHealth study sample [55]: Child scale, α=.82; Parent scale, α=.81	[55,56]
Author-developed questionnaire; acceptability and perceived impact	7 items; 5-point Likert scale; administration: web-based	One-factor scale [57]: 69% of total item variance	Internal consistency; eHealth study sample [57]: α=.94	[57]
Author-developed questionnaire; perceived impact and usability	>14 items; 5- and 10-point Likert scales; administration: not reported	Not reported	Internal consistency; eHealth study sample [58]: perceived benefits of intervention content, α=.92; perceived benefits of the interpersonal principles in the intervention, α=.85; ease of use, α=.94; ease of understanding, α=.96; ease of reading, α=.97; internal rationale, α=.96; identification/relevance, α=.96	[58,59]
Author-developed questionnaire; perceived impact, satisfaction, and usability	7 items; 4-point Likert scale; administration: not reported	Two-factor scale: utility of program (5 items), 45% of total item variance; user friendliness of program (2 items), 21% of total item variance	Internal consistency; eHealth study sample [60]: user friendliness, α=.71; utility items, α=.84	[60]
Author-developed questionnaire; acceptability, satisfaction, and perceived impact	17 items; scale type not reported; administration: not reported	Three-factor scale; eHealth study sample [61]: program evaluation (7 items); program benefits (7 items); overall satisfaction (3 items); subscales ranged from 1 (negative evaluation) to 10 (positive evaluation)	Internal consistency; eHealth study sample [61]: subscales, α=.90 or higher	[61]

[a]SUMI: Software Usability Measurement Inventory.

[b]WAMMI: Website Analysis and Measurement Inventory.

[c]TEI-SF: Treatment Evaluation Inventory- Short Form.

Delphi Consultation

Overview

In round 1, 30 researchers and 27 adolescents participated. In round 2, the number of participants decreased to 23 researchers (23/30, 77% participation rate) and 6 adolescents (6/27, 22% participation rate; Figure 1). The demographic characteristics of researcher and adolescent participants are presented in Table 4. Researcher participants were mainly women and employed in academic positions, and all participants had used a measure of user experience in their work. In addition, most adolescent participants were female participants, and most had limited experience in using eHealth interventions (despite being currently enrolled in an eHealth intervention for anxiety).

Table 4. Demographic information about the participants.

Characteristics	Round 1	Round 2
Researcher participants, n (%)	30 (100)	23 (77)
Age (years), mean (SD)	42.6 (1.4)	43.7 (1.7)
Sex, n (%)		
Female	26 (86.7)	20 (87)
Male	4 (13.3)	3 (13)
Primary role or position, n (%)		
Academic (professor and lecturer)	19 (63.3)	13 (56.5)
Scientist (researcher and research fellow)	5 (16.7)	5 (21.7)
Clinician	4 (13.3)	4 (17.4)
Trainee (PhD candidate and postdoctoral fellow)	2 (6.7)	1 (4.3)
Country, n (%)		
Australia	4 (12.9)	3 (13)
Canada	3 (9.7)	2 (8.7)
Finland	1 (3.2)	1 (4.3)
Ireland	2 (6.5)	1 (4.3)
Italy	1 (3.2)	1 (4.3)
Korea	1 (3.2)	1 (4.3)
New Zealand	4 (12.9)	3 (13)
Sweden	2 (6.5)	2 (8.7)
United Kingdom	1 (3.2)	1 (4.3)
United States	11 (35.5)	8 (34.8)
Has measured user experience, n (%)	30 (100)	23 (100)
Has developed a user experience measure, n (%)	7 (23.3)	6 (26.1)
Adolescent participants, n (%)	27 (100)	6 (22)
Age (years), mean (SD)	16.44 (0.6)	N/A[a,b]
Sex, n (%)		
Female	20 (74.1)	N/A
Male	5 (18.5)	N/A
Other	2 (7.4)	N/A
Use of eHealth programs, n (%)		
Never used until the day of the survey	14 (51.9)	N/A
<once per week	7 (25.9)	N/A
1-2 times per week	4 (14.8)	N/A
3-4 times per week	0 (0)	N/A
5-6 times per week	1 (3.7)	N/A
≥7 times per week	1 (3.7)	N/A

[a]N/A: not applicable.

[b]Demographics for adolescent participants in round 2 were not collected to ensure anonymity as per the research ethics board's requirements.

Researcher Participants

Over 2 rounds of consultation, researcher participants made several suggestions for refining the original definition of each user experience domain. The revised definitions achieved by round 2 are presented in Table 5 along with the consensus scores for the definition (the percentage of participants who *strongly agreed* or *agreed* with the definition). Researchers met or surpassed the threshold for agreement on all definitions except the definition for perceived impact. Regarding the importance

of each of the domains to the overall assessment of user experience in round 1, researchers achieved consensus in their ranking of credibility as less important to measure relative to the other 5 domains (80%; Table 6); no other rankings achieved consensus in this round. Perceived impact was ranked in both rounds as more important than the other domains, but consensus (≥80% agreement) was not achieved. Regarding when assessment of user experience should be conducted, the responses differed across domains. By round 2, researchers agreed that eHealth intervention acceptability (87% consensus) and satisfaction (97% consensus) should be measured after intervention completion. Researchers were divided on when credibility should be measured, with equal proportions of researchers indicating credibility should be measured at all 3 time points (before the intervention: 61%; during the intervention: 65%; and after the intervention: 61%).

Table 5. Working definitions of user experience domains developed with researcher participants.[a]

Domain	User experience definition	Definition consensus (%)
Acceptability	Acceptability refers to whether the intervention *content, features, and delivery* meet user *expectations* (eg, relevance, convenience, accessibility, feasibility, appropriateness, *appeal [fun, interesting, and likable], value, engaging, and privacy*). *These aspects may be different depending on the user (ie, child, adolescent, or parent).*	100
Satisfaction	Satisfaction refers to *the user's overall impression of the* intervention *and whether it meets their needs* (eg, global satisfaction rating, *value for money or time, helpful, whether they would they use it again or recommend it to a friend, and ratio between expectations and results*).	96
Credibility	Credibility refers to the extent to which the user perceives the intervention to be trustworthy and *has the potential to work* (eg, perceived accuracy and quality of information *in the intervention and/or evidence base supporting the intervention*).	96
Usability[b]	Usability refers to the user's perceived ease of use *of the intervention based on technical factors (eg, interface/equipment/reminder features or problems)* and environmental/personal factors (eg, content, time, and *competing priorities*) *that impact* the individual's use of the intervention (eg, *frequency*).	87
User-reported adherence[c]	User-reported adherence refers to how and why the user *did or did not follow* the intervention or research protocol (eg, completing outcome measures and content) as recommended.[d]	83
Perceived impact	Perceived impact refers to the extent to which the user *perceives the effect of the intervention's impacts (eg, impressions of* change in symptom levels and skills and *perception* of overall effectiveness).	78

[a]Italics represent additions or changes to the original definition. Domains are listed in descending order of consensus.

[b]Usability should also be measured in conjunction with objective measures of use (ie, intervention metadata).

[c]User-reported adherence should also be measured in conjunction with objective measures of adherence (ie, intervention metadata) and clinician expectations and adherence to the protocol (if relevant).

[d]Content removed from the definition.

Table 6. The relative rankings among researcher participants for the importance of the user experience domains across 2 rounds of consultation.

Domain	Round 1 (%)		Round 2 (%)	
	More important	Less important	More important	Less important
Acceptability	73	27	48	52
Satisfaction	43	57	30	70
Credibility	20	80	—[a]	—
Usability	47	53	26	74
User-reported adherence	47	53	26	74
Perceived impact	70	30	65	35

[a]Round 2 not conducted as consensus was achieved in round 1.

Researcher participants' opinions varied greatly on whether a universal measure of user experience was needed to enable direct comparisons between studies of eHealth interventions. By round 2, a total of 70% (16/23) stated that a universal measure was extremely or quite important, 22% (5/23) stated that it was slightly important, and 9% (2/23) stated it was slightly unimportant. In round 1, a total of 37% (11/30) of the participants also provided comments. Most stated that although a universal measure may be impractical given variability across users' developmental stage, technology types, and language, a core set of user experience items would be a valuable addition to the eHealth field. Participants suggested that other measures could be added to this core set to obtain intervention-specific feedback as needed. Moreover, 1 participant pointed out that

accepted definitions of user experience domains are also important to ensure that the domains are not used differently from study to study even if different measures are used to assess them.

Adolescent Participants

Adolescents were asked to indicate which user experience domains were important for a researcher to ask them about. In round 1, adolescents reached consensus that acceptability was more important for a researcher to ask about compared with other domains (81% consensus); satisfaction and perceived impact were indicated as important, but adolescents did not reach consensus in round 1 (78% consensus for both). After the second round, adolescents achieved consensus with satisfaction, credibility, and perceived impact (100% consensus) and usability (83% consensus) identified as important domains for researchers to ask about when measuring adolescents' user experience. After the 2 survey rounds, adolescents remained divided on the importance of measuring user-reported adherence, with 50% (3/6) of the participants rating it more important and 50% (3/6) rating it as less important

When asked to rank the domains in order of importance to one another (ie, which domains were more important than the other domains), perceived impact was considered the most important domain to measure (83% consensus) based on 2 rounds of consultation. Credibility and acceptability were also ranked highly and were considered more important than the other domains, but they did not meet or surpass the 80% threshold to achieve consensus. User-reported adherence was ranked the least important domain relative to the others (83% consensus). Usability (76%) and satisfaction (67%) were also ranked as less important to measure, although consensus was not reached for these 2 domains either.

Discussion

Summary of Principal Findings

To our knowledge, this is the first study to review how eHealth intervention studies have defined and measured user experience, consult with experts (researchers) as to how user experience could be defined and measured, and consult with users (adolescents) as to which domains they considered important for examining their user experience. In the scoping review, we made 2 important discoveries: several well-established measures are available to quantify user experiences and a large proportion of published eHealth studies did not involve the use of a well-established measure, with authors having developed user experience questions specific to their eHealth intervention. Key findings from our Delphi consultation are the alignment between researcher and adolescent relative rankings of user experience domains and the refinement of definitions for the 6 proposed domains of user experience.

Discussion of Principal Findings

We identified 10 well-established measures available in the current literature to quantify 5 of the 6 proposed domains of user experience (satisfaction, acceptability, credibility, perceived impact, and usability). Therefore, we recommend that eHealth researchers use an available, well-established measure instead of developing their own measure to assess these 5 user experience domains. This approach will allow for between-study comparisons of user experiences, including whether and how experiences with similar eHealth interventions are considered the same or different among users. Not only could such comparisons help inform parents and their children with their decisions for which eHealth intervention to use but also findings across studies could be used by researchers to identify and understand the relationship between user experience and other eHealth outcomes (ie, intervention effectiveness and program adoption or use).

To date, no well-established measure exists for studying user-reported adherence. This is not surprising, as adherence is typically measured using intervention metadata, such as the number of sessions completed and time spent per session. We propose that in addition to such objective metrics, researchers measure how and why a user *did or did not follow* the intervention or research protocol (eg, completing outcome measures and content) as recommended or as encouraged by intervention design. This subjective information can expand the understanding of metadata—for example, why did the user complete the number of sessions or outcome measures that they did. These data could be collected using a set of open-ended questions designed to be used across studies and tailored, when needed, to specific interventions, studies, or user profiles.

Definitions of the 6 domains of user experience that resulted from our international consultation with eHealth researchers offer a guidepost for new studies of user experience. Although the proposed definitions may continue to be refined over time as the eHealth field advances, we see them as an important cornerstone to user experience measurement. A set of commonly accepted domains, similar to the efforts to define, standardize, and measure eHealth adherence [16,17] and engagement [170,171], can introduce a taxonomy that can be applied across eHealth studies even if different populations, interventions, and measures are used. However, it remains unclear, as to whether a universal measure of user experience would be useful for researchers. Although 16 of the 23 of the researchers in this study responded that it is important to use a universal measure, several concerns were brought forward regarding the challenge to implementing such an approach. As a follow-up to the results we reported here, further investigation of the utility and feasibility of a core set of items and what this core set should be is needed. It is possible that consideration needs to be given to how measures can be adapted for different respondents (eg, parents, adolescents, and clinicians) or different intervention contexts (eg, open access vs therapist supported). For example, the findings of this study showed that perceived impact was rated as important by both the researchers and adolescents; however, it is entirely possible that each would describe the desired impact of the intervention in different ways. Further investigation is necessary to examine how these constructs can be best assessed using different respondents and contexts.

The Delphi consultations we conducted provide insight into the value that researchers and adolescents place on the different domains of a user's experience with an eHealth intervention. Both adolescent and researcher respondents ranked perceived impact as one of the most important aspects of a user's eHealth

experience, indicating that whether the user perceives the intervention to have had an impact on their health is a central component of user experience. There were also divergent perspectives, with one the most divergent being researchers' ranking of satisfaction among the most important domains compared with adolescents' ranking of it as less important. At the core of the definition of satisfaction is an emphasis on whether an intervention meets a user's expectations and needs. From the adolescent point of view, the definition of perceived impact, which focuses on the *impressions of change in symptom levels and skills or the perception of overall effectiveness*, may have been a more meaningful way to identify whether their needs would be met, as perceived impact directly links to an observable or tangible change in their symptoms or behavior because of the eHealth program. The ranking of credibility also differed between researchers (lower ranking) and adolescents (higher ranking). With the choice of eHealth interventions increasing, as consumers, it is not surprising that adolescents would place greater priority on credibility as an important aspect to the selection and use of an eHealth intervention. This prioritization should be considered by researchers when developing and marketing their interventions for use.

Usability was considered one of the least important domains of user experience from the perspective of both adolescent and researcher respondents. This ranking may reflect the discrepancy between the more commonly used definitions of usability and the one proposed in our Delphi consultation. Originating from the field of computer science, usability is considered a main design component (design heuristic), similar to aesthetics, user safety, or data privacy, meaning that a certain degree of (technical or functional) usability is fundamental to or expected with the use of an eHealth intervention. Although an intervention may be usable, it does not mean that it will be *used* or perceived as *useful* by adolescents [172]. Therefore, if usability is currently understood as a measure of the technical ease of use or the functionality of an intervention, then examining it may add little *richness* to the understanding of what and how users describe their user experience with eHealth intervention to be like. However, if usability comes to be associated with the conditions (barriers or constraints, facilitators, and context) of use and reasons for usefulness (meets the needs and preferences of users), then usability may be considered an important indicator of the user experience.

Future Directions

This scoping review with Delphi consultations has provided a broad overview of the current state of user experience measurement in the eHealth field along with expert (researcher) and user (adolescent) input into how user experience could be defined and measured with an eHealth intervention. An extension of this work may include investigating whether a core set of items used to measure the various domains of user experience would add value to the field and could be feasibly applied across different interventions and user populations [23]. Future research could benefit from qualitative investigations with adolescents to further define their understanding and definitions of the user experience domains within different eHealth contexts and test the feasibility of core assessment items for these domains from their perspective. It would then be beneficial to validate their definition of user experience (and the associated domains) and test the feasibility of core measurement item sets across large numbers of adolescent users of eHealth interventions internationally. With greater awareness and emphasis on patient-oriented research and improving outcomes important to patients, assessment of user experience can become an important part of patient-centered treatment planning.

Further attention could also be directed toward the definition and measurement of self-reported adherence. In our study, although a definition of self-reported adherence was achieved through consensus, its importance in measurement was not established among researcher or adolescent participants. Although a wealth of measurement studies exists for understanding adherence from an objective standpoint [16,173,174], few explanatory studies have been conducted to explore how and why the user did or did not follow the intervention and research protocol as recommended [23]. This *why* component is critical for meaningful improvement of eHealth interventions to increase program adherence and therefore achieve related health benefits [175].

Future studies may also look at how to apply both objective and subjective intervention outcomes or user experience measures to improve the validity of eHealth evaluations. For example, adolescents and researchers in our study reported perceived impact to be an important aspect of the user experience. Objective measures of intervention impact, such as changes in diagnostic severity, global functioning, symptom checklists [176], or self-reported minimal clinically important difference [177], could reinforce or complement the findings generated by more subjective measures [23]. In this way, we could better understand how various measures converge or diverge on similar user experience concepts, begin to develop more psychometrically and theoretically robust assessment measures, and establish indicators of clinically meaningful outcomes based on users' perspectives.

Limitations

Although this scoping review and associated Delphi consultations were conducted according to published guidelines, this study is not without limitations. First, our study focus was placed on eHealth interventions that were web-, computer-, or mobile-based and mediated by the internet; we excluded studies that did not primarily include these features and therefore, our results will not be representative of all technologies for which user experience may be measured. In addition, eHealth interventions being used and evaluated in health care systems that have not been scientifically investigated and reported in the published or gray literature were not included in our review. Our scoping review focus also required studies to describe the measure or measures used to collect user experience data. This requirement resulted in the exclusion of 512 studies. Given that this was a review of definition and measurement of user experience, such details were essential to understanding the current state of the eHealth field. This approach was systematic in that we applied the same working definitions to each study; however, it may have resulted in the classification of a domain of user experience that differed from what study investigators

intended. The challenge in grouping some of the studies confirms that agreement regarding definitions of user experience domains would be of value to the eHealth field. The proposed domains and definitions are not intended to be static, and we expect that they will be refined to reflect advances in the eHealth field. Finally, although we present the results from the first international Delphi consultations, our sample size was limited; particularly in representation from adolescent users.

Conclusions

eHealth interventions are now widely available for use by children, adolescents, and parents, and positive user experiences are generally reported across individual studies. The outcomes of this review and Delphi process highlight the various ways in which user experience has been defined and measured across studies, with a large proportion of research studies using study-specific, nonstandardized instruments. Through the conduct of this study, we propose definitions for 6 user experience domains: acceptability, satisfaction, credibility, usability, user-reported adherence, and perceived impact, as informed by empirical literature and agreed upon by eHealth researchers and adolescent users. Findings revealed 10 well-established measures that assess 5 of the 6 user experience domains (satisfaction, acceptability, credibility, perceived impact, and usability), and we recommend that eHealth researchers use an available, well-established measure over developing their own to assess these domains. The proposed working definitions and importance rankings from researcher and adolescent participants can be used to inform eHealth user experience research in the future and encourage consistency in reporting and to guide the development of measurement tools. Future studies should examine whether a core set of items used to measure the various domains of user experience would add value to the field and be feasibly applied across different interventions and user populations. Closing these gaps has the potential to enable comparisons across the user experience literature and better understand how user experience relates to other outcomes, such as effectiveness or objective measures of adherence, in eHealth interventions.

Acknowledgments

This work was supported by the Alberta Strategy for Support for People and Patient-Oriented Research and Trials (SUPPORT) Unit Knowledge Translation Platform, which is funded by Alberta Innovates and the Canadian Institutes of Health Research. The authors would like to acknowledge the contributions of Robin Featherstone, MLIS, who developed and conducted the initial search strategy; Tara Landry, MLIS, who peer-reviewed the search strategy and conducted an updated search; and Marcus O'Neill who assisted with article screening for eligibility and data abstraction.

Conflicts of Interest

Although SM does not own the BRAVE Program, it is possible that at some point in the future, she may receive royalties from future commercialization of the program.

Multimedia Appendix 1
Search strategy for Ovid MEDLINE Epub ahead of print, in-process, and other nonindexed citations; Ovid MEDLINE Daily; and Ovid MEDLINE.
[DOCX File , 52 KB - jmir_v23i12e25012_app1.docx]

Multimedia Appendix 2
The list of 63 test citations that were used to develop search terms for the search strategy used to identify studies on user experiences.
[DOCX File , 57 KB - jmir_v23i12e25012_app2.docx]

Multimedia Appendix 3
The definitions of user experience domains used for data extraction in the scoping review.
[DOCX File , 51 KB - jmir_v23i12e25012_app3.docx]

Multimedia Appendix 4
The parameters used for reviewing and interpreting psychometric data during the quality assessment of evaluation measures (adapted from a published table [<xref ref-type="bibr" rid="2ref25">25</xref>]).
[DOCX File , 51 KB - jmir_v23i12e25012_app4.docx]

Multimedia Appendix 5
eHealth studies that measured user experience published from 2008 to 2019.
[DOCX File , 74 KB - jmir_v23i12e25012_app5.docx]

Multimedia Appendix 6

Results from the quality assessment of user experience measures used in the 129 studies included in the review.
[DOCX File , 57 KB - jmir_v23i12e25012_app6.docx]

Multimedia Appendix 7
Evaluation measures used in the eHealth studies assessed to be promising.
[DOCX File , 54 KB - jmir_v23i12e25012_app7.docx]

References

1. Zhang T, Dong H. Human-Centred Design: An Emergent Conceptual Model. London, UK: Royal College of Art; 2009:8-10.
2. Ergonomics of human-centred system interaction. British Standards Document BS EN ISO 9241-210. 2019. URL: https://landingpage.bsigroup.com/LandingPage/Undated?UPI=000000000030388991 [accessed 2021-11-21]
3. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
4. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013 Dec;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]
5. Wilson E. Patient-Centered e-Health. Hershey, PA: IGI Global; 2009:10-25.
6. Schulze K, Kromker H. A framework to measure user experience of interactive online products. In: Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research. 2010 Presented at: 7th International Conference on Methods and Techniques in Behavioral research; August 24 - 27, 2010; Eindhoven The Netherlands p. 1-5. [doi: 10.1145/1931344.1931358]
7. 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]
8. Guagliano JM, Brown HE, Coombes E, Hughes C, Jones AP, Morton KL, et al. The development and feasibility of a randomised family-based physical activity promotion intervention: the Families Reporting Every Step to Health (FRESH) study. Pilot Feasibility Stud 2019;5:21 [FREE Full text] [doi: 10.1186/s40814-019-0408-7] [Medline: 30788135]
9. Bruggers CS, Baranowski S, Beseris M, Leonard R, Long D, Schulte E, et al. A prototype exercise-empowerment mobile video game for children with cancer, and its usability assessment: developing digital empowerment interventions for pediatric diseases. Front Pediatr 2018;6:69 [FREE Full text] [doi: 10.3389/fped.2018.00069] [Medline: 29686977]
10. De Cock N, Van Lippevelde W, Vangeel J, Notebaert M, Beullens K, Eggermont S, et al. Feasibility and impact study of a reward-based mobile application to improve adolescents' snacking habits. Public Health Nutr 2018 Aug;21(12):2329-2344. [doi: 10.1017/S1368980018000678] [Medline: 29607800]
11. Leung MM, Mateo KF, Verdaguer S, Wyka K. Testing a web-based interactive comic tool to decrease obesity risk among minority preadolescents: protocol for a pilot randomized control trial. JMIR Res Protoc 2018 Nov 09;7(11):e10682 [FREE Full text] [doi: 10.2196/10682] [Medline: 30413399]
12. March S, Spence SH, Donovan CL, Kenardy JA. Large-scale dissemination of internet-based cognitive behavioral therapy for youth anxiety: feasibility and acceptability study. J Med Internet Res 2018 Jul 04;20(7):e234. [doi: 10.2196/jmir.9211] [Medline: 29973338]
13. Turner T, Hingle M. Evaluation of a mindfulness-based mobile app aimed at promoting awareness of weight-related behaviors in adolescents: a pilot study. JMIR Res Protoc 2017 Apr 26;6(4):e67 [FREE Full text] [doi: 10.2196/resprot.6695] [Medline: 28446423]
14. Dexheimer JW, Kurowski BG, Anders SH, McClanahan N, Wade SL, Babcock L. Usability evaluation of the SMART application for youth with mTBI. Int J Med Inform 2017 Dec;97:163-170 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.007] [Medline: 27919376]
15. Feather JS, Howson M, Ritchie L, Carter PD, Parry DT, Koziol-McLain J. Evaluation methods for assessing users' psychological experiences of web-based psychosocial interventions: a systematic review. J Med Internet Res 2016;18(6):e181 [FREE Full text] [doi: 10.2196/jmir.5455] [Medline: 27363519]
16. Sieverink F, Kelders SM, van Gemert-Pijnen JE. 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]
17. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
19. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

20. 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. [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Altman DG. Practical Statistics for Medical Research. London, UK: Chapman and Hall/CRC Press; 1990:406-407.
22. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
23. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020 Feb 07;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
24. Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction to special issue: Evidence-based assessment in pediatric psychology. J Pediatr Psychol 2008 Oct;33(9):911-915 [FREE Full text] [doi: 10.1093/jpepsy/jsj115] [Medline: 16675715]
25. Phillips NM, Street M, Haesler E. A systematic review of reliable and valid tools for the measurement of patient participation in healthcare. BMJ Qual Saf 2016 Feb;25(2):110-117. [doi: 10.1136/bmjqs-2015-004357] [Medline: 26415751]
26. Hsu CC, Standford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12:1-8. [doi: 10.7275/pdz9-th90]
27. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
28. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
29. Spence SH, Donovan CL, March S, Gamble A, Anderson R, Prosser S, et al. Online CBT in the treatment of child and adolescent anxiety disorders: issues in the development of BRAVE–ONLINE and two case illustrations. Behav Cogn Psychother 2008 Jul 01;36(4):411-430. [doi: 10.1017/s135246580800444x]
30. 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]
31. Kobak KA, Mundt JC, Kennard B. Integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Ann Gen Psychiatry 2015;14:37 [FREE Full text] [doi: 10.1186/s12991-015-0077-8] [Medline: 26535048]
32. Lattie EG, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens engaged in collaborative health: the feasibility and acceptability of an online skill-building intervention for adolescents at risk for depression. Internet Interv 2017 Jun;8:15-26. [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
33. Jacobson AE, Vesely SK, Haamid F, Christian-Rancy M, O'Brien SH. Mobile application vs paper pictorial blood assessment chart to track menses in young women: a randomized cross-over design. J Pediatr Adolesc Gynecol 2018 Apr;31(2):84-88. [doi: 10.1016/j.jpag.2017.09.009] [Medline: 29030160]
34. Moulos I, Maramis C, Ioakimidis I, van den Boer J, Nolstam J, Mars M, et al. Objective and subjective meal registration via a smartphone application. In: Proceedings from the 18th International Conference on Image Analysis and Processing. 2015 Presented at: 18th International Conference on Image Analysis and Processing; September 7-11, 2015; Genoa, Italy. [doi: 10.1007/978-3-319-23222-5_50]
35. Thabrew H, Stasiak K, Merry S. Protocol for co-design, development, and open trial of a prototype game-based ehealth intervention to treat anxiety in young people with long-term physical conditions. JMIR Res Protoc 2017 Sep 22;6(9):e171 [FREE Full text] [doi: 10.2196/resprot.7250] [Medline: 28939545]
36. 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 02;7(8):e10444 [FREE Full text] [doi: 10.2196/10444] [Medline: 30072358]
37. Bernier A, Fedele D, Guo Y, Chavez S, Smith MD, Warnick J, et al. New-onset diabetes educator to educate children and their caregivers about diabetes at the time of diagnosis: usability study. JMIR Diabetes 2018 Jun 06;3(2):e10 [FREE Full text] [doi: 10.2196/diabetes.9202] [Medline: 30291069]
38. Liu A, Coleman K, Bojan K, Serrano PA, Oyedele T, Garcia A, et al. Developing a mobile app (LYNX) to support linkage to HIV/sexually transmitted infection testing and pre-exposure prophylaxis for young men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659 [FREE Full text] [doi: 10.2196/10659] [Medline: 30681964]
39. Nitsch M, Adamcik T, Kuso S, Zeiler M, Waldherr K. Usability and engagement evaluation of an unguided online program for promoting a healthy lifestyle and reducing the risk for eating disorders and obesity in the school setting. Nutrients 2019 Mar 27;11(4):713 [FREE Full text] [doi: 10.3390/nu11040713] [Medline: 30934754]
40. Madeira RN, Mestre V, Ferreirinha T. Phonological Disorders in Children? Design and user experience evaluation of a mobile serious game approach. Procedia Comput Sci 2017;113:416-421. [doi: 10.1016/j.procs.2017.08.359]

41. Bjureberg J, Sahlin H, Hedman-Lagerlöf E, Gratz KL, Tull MT, Jokinen J, et al. Extending research on Emotion Regulation Individual Therapy for Adolescents (ERITA) with nonsuicidal self-injury disorder: open pilot trial and mediation analysis of a novel online version. BMC Psychiatry 2018 Oct 11;18(1):326 [FREE Full text] [doi: 10.1186/s12888-018-1885-6] [Medline: 30305103]
42. Wilansky P, Eklund JM, Milner T, Kreindler D, Cheung A, Kovacs T, et al. Cognitive behavior therapy for anxious and depressed youth: improving homework adherence through mobile technology. JMIR Res Protoc 2016 Nov 10;5(4):e209 [FREE Full text] [doi: 10.2196/resprot.5841] [Medline: 27833071]
43. Jensen CD, Duncombe KM, Lott MA, Hunsaker SL, Duraccio KM, Woolford SJ. An evaluation of a smartphone-assisted behavioral weight control intervention for adolescents: pilot study. JMIR Mhealth Uhealth 2016 Aug 23;4(3):e102 [FREE Full text] [doi: 10.2196/mhealth.6034] [Medline: 27554704]
44. Lalouni M, Ljótsson B, Bonnert M, Hedman-Lagerlöf E, Högström J, Serlachius E, et al. Internet-delivered cognitive behavioral therapy for children with pain-related functional gastrointestinal disorders: feasibility study. JMIR Ment Health 2017 Aug 10;4(3):e32 [FREE Full text] [doi: 10.2196/mental.7985] [Medline: 28798012]
45. 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 07;6(9):e11450 [FREE Full text] [doi: 10.2196/11450] [Medline: 30194060]
46. Sun T, Dunsmuir D, Miao I, Devoy GM, West NC, Görges M, et al. In-hospital usability and feasibility evaluation of Panda, an app for the management of pain in children at home. Paediatr Anaesth 2018 Oct;28(10):897-905. [doi: 10.1111/pan.13471] [Medline: 30302882]
47. Hill RM, Pettit JW. Pilot randomized controlled trial of LEAP: a selective preventive intervention to reduce adolescents' perceived burdensomeness. J Clin Child Adolesc Psychol 2019;48(sup1):45-56. [doi: 10.1080/15374416.2016.1188705] [Medline: 27532425]
48. Law EF, Beals-Erickson SE, Noel M, Claar R, Palermo TM. Pilot randomized controlled trial of internet-delivered cognitive-behavioral treatment for pediatric headache. Headache 2015;55(10):1410-1425 [FREE Full text] [doi: 10.1111/head.12635] [Medline: 26316194]
49. Richards D, Caldwell P. Improving health outcomes sooner rather than later via an interactive website and virtual specialist. IEEE J Biomed Health Inform 2018 Sep;22(5):1699-1706. [doi: 10.1109/JBHI.2017.2782210] [Medline: 29990229]
50. Vigerland S, Ljótsson B, Thulin U, Öst L, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: a randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
51. Vigerland S, Thulin U, Ljótsson B, Svirsky L, Ost L, Lindefors N, et al. Internet-delivered CBT for children with specific phobia: a pilot study. Cogn Behav Ther 2013;42(4):303-314. [doi: 10.1080/16506073.2013.844201] [Medline: 24245708]
52. Jolstedt M, Ljótsson B, Fredlander S, Tedgård T, Hallberg A, Ekeljung A, et al. Implementation of internet-delivered CBT for children with anxiety disorders in a rural area: a feasibility trial. Internet Interv 2018 Jun;12:121-129 [FREE Full text] [doi: 10.1016/j.invent.2017.11.003] [Medline: 30135776]
53. O'Malley G, Dowdall G, Burls A, Perry IJ, Curran N. Exploring the usability of a mobile app for adolescent obesity management. JMIR Mhealth Uhealth 2014 Jun 27;2(2):e29 [FREE Full text] [doi: 10.2196/mhealth.3262] [Medline: 25098237]
54. Danielson CK, McCauley JL, Gros KS, Jones AM, Barr SC, Borkman AL, et al. SiHLEWeb.com: Development and usability testing of an evidence-based HIV prevention website for female African-American adolescents. Health Informatics J 2016 Jun;22(2):194-208 [FREE Full text] [doi: 10.1177/1460458214544048] [Medline: 25167865]
55. Palermo TM, Law EF, Fales J, Bromberg MH, Jessen-Fiddick T, Tai G. Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents: a randomized controlled multicenter trial. Pain 2016 Jan;157(1):174-185 [FREE Full text] [doi: 10.1097/j.pain.0000000000000348] [Medline: 26335910]
56. Palermo TM, Wilson AC, Peters M, Lewandowski A, Somhegyi H. Randomized controlled trial of an internet-delivered family cognitive-behavioral therapy intervention for children and adolescents with chronic pain. Pain 2009 Nov;146(1-2):205-213 [FREE Full text] [doi: 10.1016/j.pain.2009.07.034] [Medline: 19695776]
57. Carrasco AE. Acceptability of an adventure video game in the treatment of female adolescents with symptoms of depression. ResPsy 2016 Apr 18;19(1):10-18. [doi: 10.4081/ripppo.2016.182]
58. Gladstone T, Marko-Holguin M, Henry J, Fogel J, Diehl A, Van Voorhees BW. Understanding adolescent response to a technology-based depression prevention program. J Clin Child Adolesc Psychol 2014;43(1):102-114. [doi: 10.1080/15374416.2013.850697] [Medline: 24245935]
59. Gladstone TG, Marko-Holguin M, Rothberg P, Nidetz J, Diehl A, DeFrino DT, et al. An internet-based adolescent depression preventive intervention: study protocol for a randomized control trial. Trials 2015 May 01;16:203 [FREE Full text] [doi: 10.1186/s13063-015-0705-2] [Medline: 25927539]
60. Doumas DM. Web-based personalized feedback: is this an appropriate approach for reducing drinking among high school students? J Subst Abuse Treat 2015 Mar;50:76-80. [doi: 10.1016/j.jsat.2014.09.005] [Medline: 25448614]
61. Wade SL, Cassedy AE, Sklut M, Taylor HG, McNally KA, Kirkwood MW, et al. The relationship of adolescent and parent preferences for treatment modality with satisfaction, attrition, adherence, and efficacy: The Coping With Head Injury

Through Problem-Solving (CHIPS) Study. J Pediatr Psychol 2019 Apr 01;44(3):388-401. [doi: 10.1093/jpepsy/jsy087] [Medline: 30452665]
62. Bauer S, Bilić S, Reetz C, Ozer F, Becker K, Eschenbeck H, ProHEAD Consortium. Efficacy and cost-effectiveness of Internet-based selective eating disorder prevention: study protocol for a randomized controlled trial within the ProHEAD Consortium. Trials 2019 Jan 30;20(1):91 [FREE Full text] [doi: 10.1186/s13063-018-3161-y] [Medline: 30700318]
63. Nieto R, Boixadós M, Hernández E, Beneitez I, Huguet A, McGrath P. Quantitative and qualitative testing of DARWeb: an online self-guided intervention for children with functional abdominal pain and their parents. Health Informatics J 2019 Dec;25(4):1511-1527 [FREE Full text] [doi: 10.1177/1460458218779113] [Medline: 29865899]
64. Ahmed B, Monroe P, Hair A, Tan CT, Gutierrez-Osuna R, Ballard KJ. Speech-driven mobile games for speech therapy: user experiences and feasibility. Int J Speech Lang Pathol 2018 Oct 09:1-15. [doi: 10.1080/17549507.2018.1513562] [Medline: 30301384]
65. Anderson LM, Leonard S, Jonassaint J, Lunyera J, Bonner M, Shah N. Mobile health intervention for youth with sickle cell disease: impact on adherence, disease knowledge, and quality of life. Pediatr Blood Cancer 2018 Aug;65(8):e27081. [doi: 10.1002/pbc.27081] [Medline: 29693797]
66. Benyakorn S, Calub CA, Riley SJ, Schneider A, Iosif A, Solomon M, et al. Computerized cognitive training in children with autism and intellectual disabilities: feasibility and satisfaction study. JMIR Ment Health 2018 May 25;5(2):e40 [FREE Full text] [doi: 10.2196/mental.9564] [Medline: 29802090]
67. Braciszewski JM, Wernette GK, Moore RS, Bock BC, Stout RL, Chamberlain P. A pilot randomized controlled trial of a technology-based substance use intervention for youth exiting foster care. Child Youth Serv Rev 2018 Nov;94:466-476 [FREE Full text] [doi: 10.1016/j.childyouth.2018.08.011] [Medline: 31435121]
68. de la Vega R, Roset R, Galán S, Miró J. Fibroline: a mobile app for improving the quality of life of young people with fibromyalgia. J Health Psychol 2018 Jan;23(1):67-78. [doi: 10.1177/1359105316650509] [Medline: 27225285]
69. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018 Feb 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
70. Hieftje K, Pendergrass T, Montanaro E, Kyriakides T, Florsheim O, Fiellin L. "But do they like it?" Participant satisfaction and gameplay experience of a public health videogame intervention in adolescents. In: Proceedings of the IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH). 2018 Presented at: IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH); May 16-18, 2018; Vienna, Austria. [doi: 10.1109/segah.2018.8401349]
71. Husted GR, Weis J, Teilmann G, Castensøe-Seidenfaden P. Exploring the influence of a smartphone app (Young with Diabetes) on young people's self-management: qualitative study. JMIR Mhealth Uhealth 2018 Feb 28;6(2):e43 [FREE Full text] [doi: 10.2196/mhealth.8876] [Medline: 29490897]
72. Jolstedt M, Wahlund T, Lenhard F, Ljótsson B, Mataix-Cols D, Nord M, et al. Efficacy and cost-effectiveness of therapist-guided internet cognitive behavioural therapy for paediatric anxiety disorders: a single-centre, single-blind, randomised controlled trial. Lancet Child Adolesc Health 2018 Dec;2(11):792-801. [doi: 10.1016/S2352-4642(18)30275-X] [Medline: 30241993]
73. Kapitány-Fövény M, Vagdalt E, Ruttkay Z, Urbán R, Richman MJ, Demetrovics Z. Potential of an interactive drug prevention mobile phone app (Once Upon a High): questionnaire study among students. JMIR Serious Games 2018 Dec 04;6(4):e19 [FREE Full text] [doi: 10.2196/games.9944] [Medline: 30514697]
74. Klee P, Bussien C, Castellsague M, Combescure C, Dirlewanger M, Girardin C, et al. An intervention by a patient-designed do-it-yourself mobile device app reduces HbA1c in children and adolescents with type 1 diabetes: a randomized double-crossover study. Diabetes Technol Ther 2018 Dec;20(12):797-805. [doi: 10.1089/dia.2018.0255] [Medline: 30403495]
75. Leonard NR, Casarjian B, Fletcher RR, Praia C, Sherpa D, Kelemen A, et al. Theoretically-based emotion regulation strategies using a mobile app and wearable sensor among homeless adolescent mothers: acceptability and feasibility study. JMIR Pediatr Parent 2018;1(1):e1 [FREE Full text] [doi: 10.2196/pediatrics.9037] [Medline: 30637376]
76. Lorusso ML, Biffi E, Molteni M, Reni G. Exploring the learnability and usability of a near field communication-based application for semantic enrichment in children with language disorders. Assist Technol 2018;30(1):39-50. [doi: 10.1080/10400435.2016.1253046] [Medline: 28632018]
77. Parisod H, Pakarinen A, Axelin A, Löyttyniemi E, Smed J, Salanterä S. Feasibility of mobile health game "Fume" in supporting tobacco-related health literacy among early adolescents: a three-armed cluster randomized design. Int J Med Inform 2018 May;113:26-37. [doi: 10.1016/j.ijmedinf.2018.02.013] [Medline: 29602430]
78. Parisod H, Pakarinen A, Axelin A, Loyttyniemi E, Smed J, Salantera S. Acceptability of and demand for a tobacco-related health game application in early adolescents with different gender, age and gaming habits. In: Proceedings of the 1st International GamiFIN Conference. 2017 Presented at: 1st International GamiFIN Conference; May 9-10, 2017; Pori, Finland p. 135-140 URL: https://tinyurl.com/5wnan8p7
79. Ramsey RR, Holbein CE, Powers SW, Hershey AD, Kabbouche MA, O'Brien HL, et al. A pilot investigation of a mobile phone application and progressive reminder system to improve adherence to daily prevention treatment in adolescents and

young adults with migraine. Cephalalgia 2018 Dec;38(14):2035-2044. [doi: 10.1177/0333102418756864] [Medline: 29635937]
80. Verberg FL, Helmond P, Overbeek G. Study protocol: a randomized controlled trial testing the effectiveness of an online mindset intervention in adolescents with intellectual disabilities. BMC Psychiatry 2018 Dec 04;18(1):377 [FREE Full text] [doi: 10.1186/s12888-018-1939-9] [Medline: 30514245]
81. Verdaguer S, Mateo KF, Wyka K, Dennis-Tiwary TA, Leung MM. A web-based interactive tool to reduce childhood obesity risk in urban minority youth: usability testing study. JMIR Form Res 2018 Nov 01;2(2):e21 [FREE Full text] [doi: 10.2196/formative.9747] [Medline: 30684417]
82. Wade SL, Bedell G, King JA, Jacquin M, Turkstra LS, Haarbauer-Krupa J, et al. Social Participation and Navigation (SPAN) program for adolescents with acquired brain injury: pilot findings. Rehabil Psychol 2018 Aug;63(3):327-337 [FREE Full text] [doi: 10.1037/rep0000187] [Medline: 30024207]
83. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and evaluating JIApp: acceptability and usability of a smartphone app system to improve self-management in young people with juvenile idiopathic arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
84. Castensøe-Seidenfaden P, Reventlov HG, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a self-management app for young people with type 1 diabetes: methodological challenges, experiences, and recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]
85. Goyal S, Nunn CA, Rotondi M, Couperthwaite AB, Reiser S, Simone A, et al. A mobile app for the self-management of type 1 diabetes among adolescents: a randomized controlled trial. JMIR Mhealth Uhealth 2017 Jun 19;5(6):e82 [FREE Full text] [doi: 10.2196/mhealth.7336] [Medline: 28630037]
86. Hatfield M, Murray N, Ciccarelli M, Falkmer T, Falkmer M. Pilot of the BOOST-A™: an online transition planning program for adolescents with autism. Aust Occup Ther J 2017 Dec;64(6):448-456. [doi: 10.1111/1440-1630.12410] [Medline: 28722152]
87. Haug S, Castro RP, Meyer C, Filler A, Kowatsch T, Schaub MP. A mobile phone-based life skills training program for substance use prevention among adolescents: pre-post study on the acceptance and potential effectiveness of the program, Ready4life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
88. Holtz BE, Murray KM, Hershey DD, Richman J, Dunneback JK, Vyas A, et al. The design and development of MyT1DHero: a mobile app for adolescents with type 1 diabetes and their parents. J Telemed Telecare 2017 Jan 01:172-180. [doi: 10.1177/1357633X17745470] [Medline: 29228854]
89. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: a multicenter pilot clinical study. Pediatr Blood Cancer 2017 Oct;64(10):26554. [doi: 10.1002/pbc.26554] [Medline: 28423223]
90. Khalil GE, Wang H, Calabro KS, Mitra N, Shegog R, Prokhorov AV. From the experience of interactivity and entertainment to lower intention to smoke: a randomized controlled trial and path analysis of a web-based smoking prevention program for adolescents. J Med Internet Res 2017 Feb 16;19(2):e44 [FREE Full text] [doi: 10.2196/jmir.7174] [Medline: 28209560]
91. Kong G, Goldberg AL, Dallery J, Krishnan-Sarin S. An open-label pilot study of an intervention using mobile phones to deliver contingency management of tobacco abstinence to high school students. Exp Clin Psychopharmacol 2017 Oct;25(5):333-337 [FREE Full text] [doi: 10.1037/pha0000151] [Medline: 29048181]
92. Kuosmanen T, Fleming TM, Newell J, Barry MM. A pilot evaluation of the SPARX-R gaming intervention for preventing depression and improving wellbeing among adolescents in alternative education. Internet Interv 2017 Jun;8:40-47 [FREE Full text] [doi: 10.1016/j.invent.2017.03.004] [Medline: 30135827]
93. Lee J, Song S, Ahn JS, Kim Y, Lee JE. Use of a mobile application for self-monitoring dietary intake: feasibility test and an intervention study. Nutrients 2017 Jul 13;9(7):13 [FREE Full text] [doi: 10.3390/nu9070748] [Medline: 28703758]
94. Lyles AA, Amresh A, Huberty J, Todd M, Lee RE. A mobile, avatar-based app for improving body perceptions among adolescents: a pilot test. JMIR Serious Games 2017 Mar 02;5(1):e4 [FREE Full text] [doi: 10.2196/games.6354] [Medline: 28254737]
95. McManama O'Brien KH, LeCloux M, Ross A, Gironda C, Wharff EA. A pilot study of the acceptability and usability of a smartphone application intervention for suicidal adolescents and their parents. Arch Suicide Res 2016 May 2:1-11. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
96. Narad ME, Bedell G, King JA, Johnson J, Turkstra LS, Haarbauer-Krupa J, et al. Social Participation and Navigation (SPAN): description and usability of app-based coaching intervention for adolescents with TBI. Dev Neurorehabil 2017 Aug 01:1-10. [doi: 10.1080/17518423.2017.1354092] [Medline: 28762859]
97. Navarro J, Escobar P, Cebolla A, Lison J, Pitti J, Guizerres J, et al. Exergames on line for childhood obesity: using a web platform as an ambulatory program to increase the acceptance and adherence to physical activity (PA). In: Proceedings of the International Summit on eHealth - eHealth 360°. 2016 Presented at: International Summit on eHealth - eHealth 360°; June 14-16, 2016; Budapest, Hungary p. 127-134. [doi: 10.1007/978-3-319-49655-9_17]
98. Newton AS, Dow N, Dong K, Fitzpatrick E, Wild TC, Johnson DW, Pediatric Emergency Research Canada. A randomised controlled pilot trial evaluating feasibility and acceptability of a computer-based tool to identify and reduce harmful and

hazardous drinking among adolescents with alcohol-related presentations in Canadian pediatric emergency departments. BMJ Open 2017 Aug 11;7(8):e015423 [FREE Full text] [doi: 10.1136/bmjopen-2016-015423] [Medline: 28801399]
99. Stoll R, Pina A, Gary K, Amresh A. Usability of a smartphone application to support the prevention and early intervention of anxiety in youth. Cogn Behav Pract 2017 Nov;24(4):393-404 [FREE Full text] [doi: 10.1016/j.cbpra.2016.11.002] [Medline: 29056845]
100. Widman L, Golin CE, Kamke K, Massey J, Prinstein MJ. Feasibility and acceptability of a web-based HIV/STD prevention program for adolescent girls targeting sexual communication skills. Health Educ Res 2017 Aug 01;32(4):343-352 [FREE Full text] [doi: 10.1093/her/cyx048] [Medline: 28854574]
101. Chapman R, Loades M, O'Reilly G, Coyle D, Patterson M, Salkovskis P. 'Pesky gNATs': investigating the feasibility of a novel computerized CBT intervention for adolescents with anxiety and/or depression in a Tier 3 CAMHS setting. Cogn Behav Ther 2016 Dec 01;9:e35. [doi: 10.1017/s1754470x16000222]
102. Glynn P, Eom S, Zelko F, Koh S. Feasibility of a mobile cognitive intervention in childhood absence epilepsy. Front Hum Neurosci 2016;10:575 [FREE Full text] [doi: 10.3389/fnhum.2016.00575] [Medline: 27895568]
103. Kurowski BG, Wade SL, Dexheimer JW, Dyas J, Zhang N, Babcock L. Feasibility and potential benefits of a web-based intervention delivered acutely after mild traumatic brain injury in adolescents: a pilot study. J Head Trauma Rehabil 2016;31(6):369-378 [FREE Full text] [doi: 10.1097/HTR.0000000000000180] [Medline: 26360000]
104. Svensson A, Magnusson M, Larsson C. Overcoming barriers: adolescents' experiences using a mobile phone dietary assessment app. JMIR Mhealth Uhealth 2016 Jul 29;4(3):e92 [FREE Full text] [doi: 10.2196/mhealth.5700] [Medline: 27473462]
105. Armbrust W, Bos JJ, Cappon J, van Rossum MA, Sauer PJ, Wulffraat N, et al. Design and acceptance of Rheumates@Work, a combined internet-based and in person instruction model, an interactive, educational, and cognitive behavioral program for children with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2015 Jul 23;13:31 [FREE Full text] [doi: 10.1186/s12969-015-0029-5] [Medline: 26202161]
106. Blackman KC, Zoellner J, Kadir A, Dockery B, Johnson SB, Almeida FA, et al. Examining the feasibility of smartphone game applications for physical activity promotion in middle school students. Games Health J 2015 Oct;4(5):409-419. [doi: 10.1089/g4h.2014.0120] [Medline: 26287931]
107. Brady SS, Sieving RE, Terveen LG, Rosser BR, Kodet AJ, Rothberg VD. An interactive website to reduce sexual risk behavior: process evaluation of TeensTalkHealth. JMIR Res Protoc 2015 Sep 02;4(3):e106 [FREE Full text] [doi: 10.2196/resprot.3440] [Medline: 26336157]
108. Bul KC, Franken IH, Van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, et al. Development and user satisfaction of "Plan-It Commander," a serious game for children with ADHD. Games Health J 2015 Dec;4(6):502-512. [doi: 10.1089/g4h.2015.0021] [Medline: 26325247]
109. Cox LE, Ashford JM, Clark KN, Martin-Elbahesh K, Hardy KK, Merchant TE, et al. Feasibility and acceptability of a remotely administered computerized intervention to address cognitive late effects among childhood cancer survivors. Neurooncol Pract 2015 Jun;2(2):78-87 [FREE Full text] [doi: 10.1093/nop/npu036] [Medline: 27054041]
110. Enah C, Piper K, Moneyham L. Qualitative evaluation of the relevance and acceptability of a web-based HIV prevention game for rural adolescents. J Pediatr Nurs 2015 Apr;30(2):321-328. [doi: 10.1016/j.pedn.2014.09.004] [Medline: 25245160]
111. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
112. Kobak KA, Mundt JC, Kennard B. Erratum to: integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Ann Gen Psychiatry 2016;15:2 [FREE Full text] [doi: 10.1186/s12991-015-0089-4] [Medline: 26779277]
113. Marsac ML, Winston FK, Hildenbrand AK, Kohser KL, March S, Kenardy J, et al. Systematic, theoretically-grounded development and feasibility testing of an innovative, preventive web-based game for children exposed to acute trauma. Clin Pract Pediatr Psychol 2015;3(1):12-24 [FREE Full text] [doi: 10.1037/cpp0000080] [Medline: 25844276]
114. Marsch LA, Guarino H, Grabinski MJ, Syckes C, Dillingham ET, Xie H, et al. Comparative effectiveness of web-based vs. educator-delivered HIV prevention for adolescent substance users: a randomized controlled trial. J Subst Abuse Treat 2015 Dec;59:30-37 [FREE Full text] [doi: 10.1016/j.jsat.2015.07.003] [Medline: 26293644]
115. 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]
116. Pretlow RA, Stock CM, Allison S, Roeger L. Treatment of child/adolescent obesity using the addiction model: a smartphone app pilot study. Child Obes 2015 Jun;11(3):248-259 [FREE Full text] [doi: 10.1089/chi.2014.0124] [Medline: 25760813]
117. Sousa P, Fonseca H, Gaspar P, Gaspar F. Controlled trial of an internet-based intervention for overweight teens (Next.Step): effectiveness analysis. Eur J Pediatr 2015 Sep;174(9):1143-1157 [FREE Full text] [doi: 10.1007/s00431-015-2502-z] [Medline: 25772743]
118. Sousa P, Fonseca H, Gaspar P, Gaspar F. Usability of an internet-based platform (Next.Step) for adolescent weight management. J Pediatr (Rio J) 2015;91(1):68-74 [FREE Full text] [doi: 10.1016/j.jped.2014.05.010] [Medline: 25245364]

119. Sze YY, Daniel TO, Kilanowski CK, Collins RL, Epstein LH. Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: a feasibility study. JMIR Mhealth Uhealth 2015;3(4):e97 [FREE Full text] [doi: 10.2196/mhealth.4603] [Medline: 26678959]
120. Voerman JS, Remerie S, Westendorp T, Timman R, Busschbach JJ, Passchier J, et al. Effects of a guided internet-delivered self-help intervention for adolescents with chronic pain. J Pain 2015 Nov;16(11):1115-1126. [doi: 10.1016/j.jpain.2015.07.011] [Medline: 26281947]
121. Williamson H, Griffiths C, Harcourt D. Developing young person's Face IT: online psychosocial support for adolescents struggling with conditions or injuries affecting their appearance. Health Psychol Open 2015 Jul;2(2):2055102915619092 [FREE Full text] [doi: 10.1177/2055102915619092] [Medline: 28070380]
122. Williamson H, Hamlet C, White P, Marques EMR, Cadogan J, Perera R, et al. Study protocol of the YP Face IT feasibility study: comparing an online psychosocial intervention versus treatment as usual for adolescents distressed by appearance-altering conditions/injuries. BMJ Open 2016 Dec 03;6(10):e012423 [FREE Full text] [doi: 10.1136/bmjopen-2016-012423] [Medline: 27697878]
123. Wozney L, Baxter P, Newton AS. Usability evaluation with mental health professionals and young people to develop an internet-based cognitive-behaviour therapy program for adolescents with anxiety disorders. BMC Pediatr 2015;15:213 [FREE Full text] [doi: 10.1186/s12887-015-0534-1] [Medline: 26675420]
124. Bannink R, Broeren S, van Zwanenburg E, van As E, van de Looij-Jansen P, Raat H. Use and appreciation of a web-based, tailored intervention (E-health4Uth) combined with counseling to promote adolescents' health in preventive youth health care: survey and log-file analysis. JMIR Res Protoc 2014 Jan 06;3(1):e3 [FREE Full text] [doi: 10.2196/resprot.2855] [Medline: 24394676]
125. Breakey VR, Ignas DM, Warias AV, White M, Blanchette VS, Stinson JN. A pilot randomized control trial to evaluate the feasibility of an internet-based self-management and transitional care program for youth with haemophilia. Haemophilia 2014 Nov;20(6):784-793. [doi: 10.1111/hae.12488] [Medline: 25311370]
126. Comer JS, Furr JM, Cooper-Vince CE, Kerns CE, Chan PT, Edson AL, et al. Internet-delivered, family-based treatment for early-onset OCD: a preliminary case series. J Clin Child Adolesc Psychol 2014;43(1):74-87 [FREE Full text] [doi: 10.1080/15374416.2013.855127] [Medline: 24295036]
127. Hetrick S, Yuen HP, Cox G, Bendall S, Yung A, Pirkis J, et al. Does cognitive behavioural therapy have a role in improving problem solving and coping in adolescents with suicidal ideation? Cogn Behav Therap 2014 Aug 07;7:e13. [doi: 10.1017/s1754470x14000129]
128. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]
129. Manicavasagar V, Horswood D, Burckhardt R, Lum A, Hadzi-Pavlovic D, Parker G. Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. J Med Internet Res 2014;16(6):e140 [FREE Full text] [doi: 10.2196/jmir.3176] [Medline: 24901900]
130. Newton NC, Conrod PJ, Rodriguez DM, Teesson M. A pilot study of an online universal school-based intervention to prevent alcohol and cannabis use in the UK. BMJ Open 2014 May 19;4(5):e004750 [FREE Full text] [doi: 10.1136/bmjopen-2013-004750] [Medline: 24840248]
131. Ybarra ML, Bull SS, Prescott TL, Birungi R. Acceptability and feasibility of CyberSenga: an internet-based HIV-prevention program for adolescents in Mbarara, Uganda. AIDS Care 2014 Apr;26(4):441-447 [FREE Full text] [doi: 10.1080/09540121.2013.841837] [Medline: 24093828]
132. Ezendam NP, Noordegraaf VS, Kroeze W, Brug J, Oenema A. Process evaluation of FATaintPHAT, a computer-tailored intervention to prevent excessive weight gain among Dutch adolescents. Health Promot Int 2013 Mar;28(1):26-35. [doi: 10.1093/heapro/das021] [Medline: 22692481]
133. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
134. Hardy KK, Willard VW, Allen TM, Bonner MJ. Working memory training in survivors of pediatric cancer: a randomized pilot study. Psychooncology 2013 Aug;22(8):1856-1865 [FREE Full text] [doi: 10.1002/pon.3222] [Medline: 23203754]
135. Harrell W, Eack S, Hooper SR, Keshavan MS, Bonner MS, Schoch K, et al. Feasibility and preliminary efficacy data from a computerized cognitive intervention in children with chromosome 22q11.2 deletion syndrome. Res Dev Disabil 2013 Sep;34(9):2606-2613 [FREE Full text] [doi: 10.1016/j.ridd.2013.05.009] [Medline: 23751300]
136. Ritterband LM, Thorndike FP, Lord HR, Borowitz S, Walker LS, Ingersoll KS, et al. An RCT of an internet intervention for pediatric encopresis with one year follow-up. Clin Pract Pediatr Psychol 2013 Mar;1(1):68-80 [FREE Full text] [doi: 10.1037/cpp0000007] [Medline: 24040581]
137. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]

138. Bradley KL, Robinson LM, Brannen CL. Adolescent help-seeking for psychological distress, depression, and anxiety using an Internet program. Int J Ment Health Prom 2012 Feb;14(1):23-34. [doi: 10.1080/14623730.2012.665337]
139. Lau EY, Lau PW, Chung P, Ransdell LB, Archer E. Evaluation of an internet-short message service-based intervention for promoting physical activity in Hong Kong Chinese adolescent school children: a pilot study. Cyberpsychol Behav Soc Netw 2012 Aug;15(8):425-434. [doi: 10.1089/cyber.2012.0161] [Medline: 22897473]
140. Shegog R, Markham CM, Leonard AD, Bui TC, Paul ME. "+CLICK": pilot of a web-based training program to enhance ART adherence among HIV-positive youth. AIDS Care 2012;24(3):310-318. [doi: 10.1080/09540121.2011.608788] [Medline: 21899488]
141. Whittaker R, Merry S, Stasiak K, McDowell H, Doherty I, Shepherd M, et al. MEMO--a mobile phone depression prevention intervention for adolescents: development process and postprogram findings on acceptability from a randomized controlled trial. J Med Internet Res 2012;14(1):e13 [FREE Full text] [doi: 10.2196/jmir.1857] [Medline: 22278284]
142. Coyle D, McGlade N, Doherty G, O'Reilly G. Exploratory evaluations of a computer game supporting cognitive behavioural therapy for adolescents. In: Proceedings of the 29th Annual CHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; Vancouver BC Canada p. 2937-2946. [doi: 10.1145/1978942.1979378]
143. Iloabachie C, Wells C, Goodwin B, Baldwin M, Vanderplough-Booth K, Gladstone T, et al. Adolescent and parent experiences with a primary care/Internet-based depression prevention intervention (CATCH-IT). Gen Hosp Psychiatry 2011;33(6):543-555. [doi: 10.1016/j.genhosppsych.2011.08.004] [Medline: 21958447]
144. Marsch LA, Grabinski MJ, Bickel WK, Desrosiers A, Guarino H, Muehlbach B, et al. Computer-assisted HIV prevention for youth with substance use disorders. Subst Use Misuse 2011;46(1):46-56 [FREE Full text] [doi: 10.3109/10826084.2011.521088] [Medline: 21190405]
145. Tillfors M, Andersson G, Ekselius L, Furmark T, Lewenhaupt S, Karlsson A, et al. A randomized trial of internet-delivered treatment for social anxiety disorder in high school students. Cogn Behav Ther 2011;40(2):147-157. [doi: 10.1080/16506073.2011.555486] [Medline: 25155815]
146. Newton NC, Teesson M, Vogl LE, Andrews G. Internet-based prevention for alcohol and cannabis use: final results of the Climate Schools course. Addiction 2010 Apr;105(4):749-759. [doi: 10.1111/j.1360-0443.2009.02853.x] [Medline: 20148791]
147. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
148. Markham CM, Shegog R, Leonard AD, Bui TC, Paul ME. +CLICK: harnessing web-based training to reduce secondary transmission among HIV-positive youth. AIDS Care 2009 May;21(5):622-631 [FREE Full text] [doi: 10.1080/09540120802385637] [Medline: 19444671]
149. Landback J, Prochaska M, Ellis J, Dmochowska K, Kuwabara SA, Gladstone T, et al. From prototype to product: development of a primary care/internet based depression prevention intervention for adolescents (CATCH-IT). Community Ment Health J 2009 Oct;45(5):349-354. [doi: 10.1007/s10597-009-9226-3] [Medline: 19641992]
150. Long AC, Palermo TM. Brief report: web-based management of adolescent chronic pain: development and usability testing of an online family cognitive behavioral therapy program. J Pediatr Psychol 2009 Jun;34(5):511-516 [FREE Full text] [doi: 10.1093/jpepsy/jsn082] [Medline: 18669578]
151. O'Conner-Von S. Preparation of adolescents for outpatient surgery: using an internet program. AORN J 2008 Feb;87(2):374-398. [doi: 10.1016/j.aorn.2007.07.024] [Medline: 18262002]
152. Ritterband LM, Ardalan K, Thorndike FP, Magee JC, Saylor DK, Cox DJ, et al. Real world use of an internet intervention for pediatric encopresis. J Med Internet Res 2008 Jun 30;10(2):e16 [FREE Full text] [doi: 10.2196/jmir.1081] [Medline: 18653440]
153. World Health Organization. Young people's health--a challenge for society. Report of a WHO Study Group on young people and "Health for All by the Year 2000". World Health Organ Tech Rep Ser 1986;731:1-117. [Medline: 3085358]
154. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int Hum Comput Interact 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
155. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comp Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
156. Lewis J, Sauro J. The factor structure of the system usability scale. In: Proceedings of the International Conference on Human Centered Design. 2009 Presented at: International Conference on Human Centered Design; July 19-24, 2009; San Diego, CA, USA p. 94-103. [doi: 10.1007/978-3-642-02806-9_12]
157. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
158. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese validation of the System Usability Scale (SUS). Procedia Comput Sci 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]
159. Greenfield T, Attkisson C. The Service Satisfaction Scale-30. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcome Assessment. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:799-811.

160. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
161. Johnson D, Gardner MJ, Perry R. Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). Int J Hum Comput Stud 2018 Oct;118:38-46. [doi: 10.1016/j.ijhcs.2018.05.003]
162. Poels K, de Kort YA, IJsselsteijn WA. D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games. In: European Community - New and Emerging Science and Technology (NEST) Programme. Eindhoven: Technische Universiteit Eindhoven; 2007:1-46.
163. Athay MM, Bickman L. Development and psychometric evaluation of the youth and caregiver Service Satisfaction Scale. Adm Policy Ment Health 2012 Mar;39(1-2):71-77 [FREE Full text] [doi: 10.1007/s10488-012-0407-y] [Medline: 22407558]
164. Kelley ML, Heffer RW, Gresham FM, Elliott SN. Development of a modified treatment evaluation inventory. J Psychopathol Behav Assess 1989 Sep;11(3):235-247. [doi: 10.1007/bf00960495]
165. Newton JT, Sturmey P. Development of a short form of the Treatment Evaluation Inventory for acceptability of psychological interventions. Psychol Rep 2004 Apr;94(2):475-481. [doi: 10.2466/pr0.94.2.475-481] [Medline: 15154175]
166. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for reliability and validity. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2018 Sep 27;62(1):1414-1418. [doi: 10.1177/1541931218621322]
167. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychother Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
168. Kirakowski J. The use of questionnaire methods for usability assessment. Background Notes on the SUMI Questionnaire. 1994. URL: https://www.researchgate.net/profile/Jurek_Kirakowski/project/SUMI-Questionnaire/attachment/5ae302e1b53d2f63c3c86f11/AS:619982895456256@1524826849051/download/SUMIPAPP.pdf?context=ProjectUpdatesLog [accessed 2021-11-21]
169. Kirakowski J, Claridge N, Whitehand R. Human centered measures of success in web site design. In: Proceedings of the 4th Conference on Human Factors & the Web. 1998 Presented at: 4th Conference on Human Factors & the Web; June 5, 1998; Basking Ridge, NJ, USA URL: https://scholar.google.com/scholar_lookup?hl=en&publication_year=1998&author=J.+Kirakowski&author=N+Claridge&author=R+Whitehand&title=Human+Centered+Measures+of+Success+in+Web+Site+Design
170. 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]
171. 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]
172. Baumel A, Muench F. Heuristic evaluation of eHealth interventions: establishing standards that relate to the therapeutic process perspective. JMIR Ment Health 2016 Jan 13;3(1):e5 [FREE Full text] [doi: 10.2196/mental.4563] [Medline: 26764209]
173. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
174. Arsenijevic J, Tummers L, Bosma N. Adherence to electronic health tools among vulnerable groups: systematic literature review and meta-analysis. J Med Internet Res 2020 Feb 06;22(2):e11613 [FREE Full text] [doi: 10.2196/11613] [Medline: 32027311]
175. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
176. Stinson J, Wilson R, Gill N, Yamada J, Holt J. A systematic review of internet-based self-management interventions for youth with health conditions. J Pediatr Psychol 2009 Jun;34(5):495-510 [FREE Full text] [doi: 10.1093/jpepsy/jsn115] [Medline: 19029142]
177. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989 Dec;10(4):407-415. [doi: 10.1016/0197-2456(89)90005-6] [Medline: 2691207]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
REDCap: Research Electronic Data Capture

Edited by R Kukafka; submitted 14.10.20; peer-reviewed by B Holtz, K Mateo, B Sindelar, M Bestek; comments to author 13.03.21; revised version received 27.03.21; accepted 23.09.21; published 02.12.21.

Please cite as:
Newton AS, March S, Gehring ND, Rowe AK, Radomski AD
Establishing a Working Definition of User Experience for eHealth Interventions of Self-reported User Experience Measures With eHealth Researchers and Adolescents: Scoping Review
J Med Internet Res 2021;23(12):e25012
URL: https://www.jmir.org/2021/12/e25012
doi:10.2196/25012
PMID:34860671

Review

Early Patient-Centered Outcomes Research Experience With the Use of Telehealth to Address Disparities: Scoping Review

James E Bailey[1], MD, MPH; Cathy Gurgol[2], MS; Eric Pan[3], MSc, MD; Shirilyn Njie[3], MSc; Susan Emmett[4], MD, MPH; Justin Gatwood[5], MPH, PhD; Lynne Gauthier[6], PhD; Lisa G Rosas[7,8], MPH, PhD; Shannon M Kearney[9], MPH, DrPH; Samantha Kleindienst Robler[10], PhD; Raymona H Lawrence[11], MPH, DrPH; Karen L Margolis[12], MD, MPH; Ifeyinwa Osunkwo[13], MD, MPH; Denise Wilfley[14], PhD; Vallabh O Shah[15], MSc, PhD

[1]Tennessee Population Health Consortium, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Patient-Centered Outcomes Research Institute, Washington, DC, United States

[3]Westat Inc, Center for Healthcare Delivery Research and Evaluation, Rockville, MD, United States

[4]Department of Head and Neck Surgery and Communication Sciences, Duke University School of Medicine, Duke Global Health Institute, Durham, NC, United States

[5]College of Pharmacy, University of Tennessee Health Science Center, Memphis, TN, United States

[6]Department of Physical Therapy and Kinesiology, Zuckerberg College of Health Sciences, University of Massachusetts, Lowell, MA, United States

[7]Department of Epidemiology and Population Health, Division of Primary Care and Population Health, Stanford School of Medicine, Palo Alto, CA, United States

[8]Department of Medicine, Division of Primary Care and Population Health, Stanford School of Medicine, Palo Alto, CA, United States

[9]Solution Insights & Validation, Highmark Health, Pittsburgh, PA, United States

[10]Norton Sound Health Corporation, Nome, AK, United States

[11]Community Health Behavior and Education, Jiann-Ping College of Public Health, Georgia Southern University, Statesboro, GA, United States

[12]HealthPartners Institute, Minneapolis, MN, United States

[13]Cancer Care, Levine Cancer Institute, Atrium Health, Charlotte, NC, United States

[14]Department of Psychiatry, College of Medicine, Washington University in St. Louis, St Louis, MO, United States

[15]Department of Internal Medicine and Biochemistry, School of Medicine, University of New Mexico, Albuquerque, NM, United States

Corresponding Author:
James E Bailey, MD, MPH
Tennessee Population Health Consortium
University of Tennessee Health Science Center
Coleman Building, Suite D222
956 Court Avenue
Memphis, TN, 38163
United States
Phone: 1 9014482475
Email: jeb@uthsc.edu

Abstract

Background: Health systems and providers across America are increasingly employing telehealth technologies to better serve medically underserved low-income, minority, and rural populations at the highest risk for health disparities. The Patient-Centered Outcomes Research Institute (PCORI) has invested US $386 million in comparative effectiveness research in telehealth, yet little is known about the key early lessons garnered from this research regarding the best practices in using telehealth to address disparities.

Objective: This paper describes preliminary lessons from the body of research using study findings and case studies drawn from PCORI seminal patient-centered outcomes research (PCOR) initiatives. The primary purpose was to identify common barriers and facilitators to implementing telehealth technologies in populations at risk for disparities.

Methods: A systematic scoping review of telehealth studies addressing disparities was performed. It was guided by the Arksey and O'Malley Scoping Review Framework and focused on PCORI's active portfolio of telehealth studies and key PCOR identified by study investigators. We drew on this broad literature using illustrative examples from early PCOR experience and published literature to assess barriers and facilitators to implementing telehealth in populations at risk for disparities, using the active

implementation framework to extract data. Major themes regarding how telehealth interventions can overcome barriers to telehealth adoption and implementation were identified through this review using an iterative Delphi process to achieve consensus among the PCORI investigators participating in the study.

Results: PCORI has funded 89 comparative effectiveness studies in telehealth, of which 41 assessed the use of telehealth to improve outcomes for populations at risk for health disparities. These 41 studies employed various overlapping modalities including mobile devices (29/41, 71%), web-based interventions (30/41, 73%), real-time videoconferencing (15/41, 37%), remote patient monitoring (8/41, 20%), and store-and-forward (ie, asynchronous electronic transmission) interventions (4/41, 10%). The studies targeted one or more of PCORI's priority populations, including racial and ethnic minorities (31/41, 41%), people living in rural areas, and those with low income/low socioeconomic status, low health literacy, or disabilities. Major themes identified across these studies included the importance of patient-centered design, cultural tailoring of telehealth solutions, delivering telehealth through trusted intermediaries, partnering with payers to expand telehealth reimbursement, and ensuring confidential sharing of private information.

Conclusions: Early PCOR evidence suggests that the most effective health system- and provider-level telehealth implementation solutions to address disparities employ patient-centered and culturally tailored telehealth solutions whose development is actively guided by the patients themselves to meet the needs of specific communities and populations. Further, this evidence shows that the best practices in telehealth implementation include delivery of telehealth through trusted intermediaries, close partnership with payers to facilitate reimbursement and sustainability, and safeguards to ensure patient-guided confidential sharing of personal health information.

(J Med Internet Res 2021;23(12):e28503) doi:10.2196/28503

KEYWORDS

telehealth; scoping review; disparities; implementation science

Introduction

Background

People living in medically underserved, low-income, or rural areas; those from racial and ethnic minorities; members of the lesbian, gay, bisexual, and transgender (LGBT) community; and those with limited English proficiency or disabilities often face substantial barriers to accessing needed health care and are at high risk for health disparities [1,2]. About one-quarter of the US population experiences geographic and/or economic disparities. For example, 34 million Americans (10.5%) live in poverty [3] and some 46 million Americans (15%) live in rural areas [4]. These populations are more likely to report issues with access to primary and specialty care, longer travel time to providers, lower rates of insurance coverage, and higher rates of chronic conditions than urban resident [2,5-7]. Racial and ethnic disparities are also common and are strongly linked to social determinants of health [1,5,8].

Other populations at particularly high risk for disparities include low-income African Americans in medically underserved urban and rural areas of the South [9-11], American Indian populations in rural areas [8,12,13], the mostly Alaskan Native residents of Alaska [14], and Latinx Americans, who face language, insurance, and other barriers to access to care [8]. Additional groups identified by the Patient-Centered Outcomes Research Institute (PCORI) as priority populations include LGBT persons, those with low health literacy/numeracy and limited English proficiency, and those with disabilities [15-21]. All these populations face barriers related to the social determinants of health that prevent adequate access to care [13,19-23], resulting in poor health outcomes across a multitude of domains.

Potential for Using Telehealth to Address Disparities

Telehealth modalities have great potential to help overcome geographic, socioeconomic, cultural, and language barriers related to the social determinants of health and enhance access to essential health services for high-risk populations [24,25]. Telehealth has traditionally been a way to provide health care access in rural communities, and its adoption has accelerated for these populations during the COVID-19 pandemic [26,27]. While the potential for telehealth to address disparities, such as transportation and availability of providers, is high, preliminary data during the COVID-19 pandemic suggest that differences in internet and telehealth access may actually compound disparities in chronic disease outcomes [28]. Telehealth can be used for multiple purposes, including direct delivery of care to patients and their caregivers, remote monitoring of health outcomes, education of patients and caregivers, support for self-care and health behavior change, assistance with health care decision-making, and communication of test results [29]. Likewise, telehealth can employ multiple communication modalities, including telephone, video teleconference, text messaging, mobile apps, and wearable monitors, that allow providers to remotely communicate, monitor, and share information with patients [30-32]. Previous research has documented the effectiveness of these telehealth modalities in reaching rural populations to provide care, prescreening evaluations, and patient education [25,33,34], yet data on the effectiveness of telehealth in other populations is more limited [29-32]. Moreover, while many health care delivery systems have implemented telehealth solutions to address barriers to access and better serve populations at risk for disparities, outcomes and sustainability have been variable [24,25,35].

Study Aims

The primary purpose of this scoping review was to identify key barriers to telehealth implementation and describe how barriers can be addressed in populations at risk for disparities. PCORI has invested US $386 million in comparative effectiveness research in telehealth, yet little is known about the key early lessons of this research for using telehealth to address disparities. This review describes preliminary lessons from early patient-centered outcomes research (PCOR) experience and literature, using study findings and illustrative case studies drawn from these seminal PCOR initiatives and literature.

Methods

Overview

A systematic scoping review of telehealth studies addressing disparities was performed and was guided by the Arksey and O'Malley Scoping Review Framework [36], with a focus on PCORI's active portfolio of telehealth studies and key supporting PCOR identified by several of the investigators.

Identifying the Research Question

The purpose of this scoping review was to describe barriers and facilitators to implementing telehealth or remote interventions in populations at risk of health disparities (eg, racial and ethnic minority groups or people living in rural areas) within comparative effectiveness research projects funded by PCORI. The specific question addressed was as follows: "How have PCORI-funded investigators overcome barriers to implementing telehealth interventions in populations at risk for disparities?"

Identifying Relevant Studies

Between December 2012 and March 2019, PCORI funded 84 comparative effective research projects, in which at least one of the study's comparison arms used telehealth to improve health outcomes. PCORI defines telehealth as the delivery of health services via remote telecommunication modalities, such as telephonic communication, remote monitoring devices, real-time videoconferencing, and mobile devices. For this scoping review, PCORI staff first reviewed the telehealth portfolio to identify studies aiming to improve health outcomes for populations at risk of or facing disparities. Studies included those that targeted at least one of the following PCORI priority populations at risk for disparities: people living in medically underserved, low-income, or rural areas; those representing racial and ethnic minorities; members of the LGBT community; those with low health literacy/numeracy and limited English proficiency; and those with disabilities [37]. Other PCORI populations of interest, including older adults, women, children, individuals with multiple chronic diseases, individuals with rare diseases, individuals whose genetic makeup affects their medical outcomes, veterans, and members of the armed forces and their families, were not specifically included in this scoping review.

Of the 84 PCORI-funded studies in this portfolio, PCORI staff identified 41 studies focused on improving health outcomes for populations at risk of or facing disparities by confirming the target population for each study. Qualifying studies were further reviewed by PCORI staff to determine if any barriers to implementation of the planned telehealth intervention were experienced and whether the investigators were able to mitigate these barriers to successfully implement the study telehealth intervention. A convenience sample of 8 studies was selected based on their representativeness (ie, to ensure that all major targeted populations at risk for disparities were included), study phase (ie, studies either completed or nearing completion), and perceived principal investigator availability to provide illustrative examples of successful approaches for working through obstacles in implementing telehealth interventions for diverse populations.

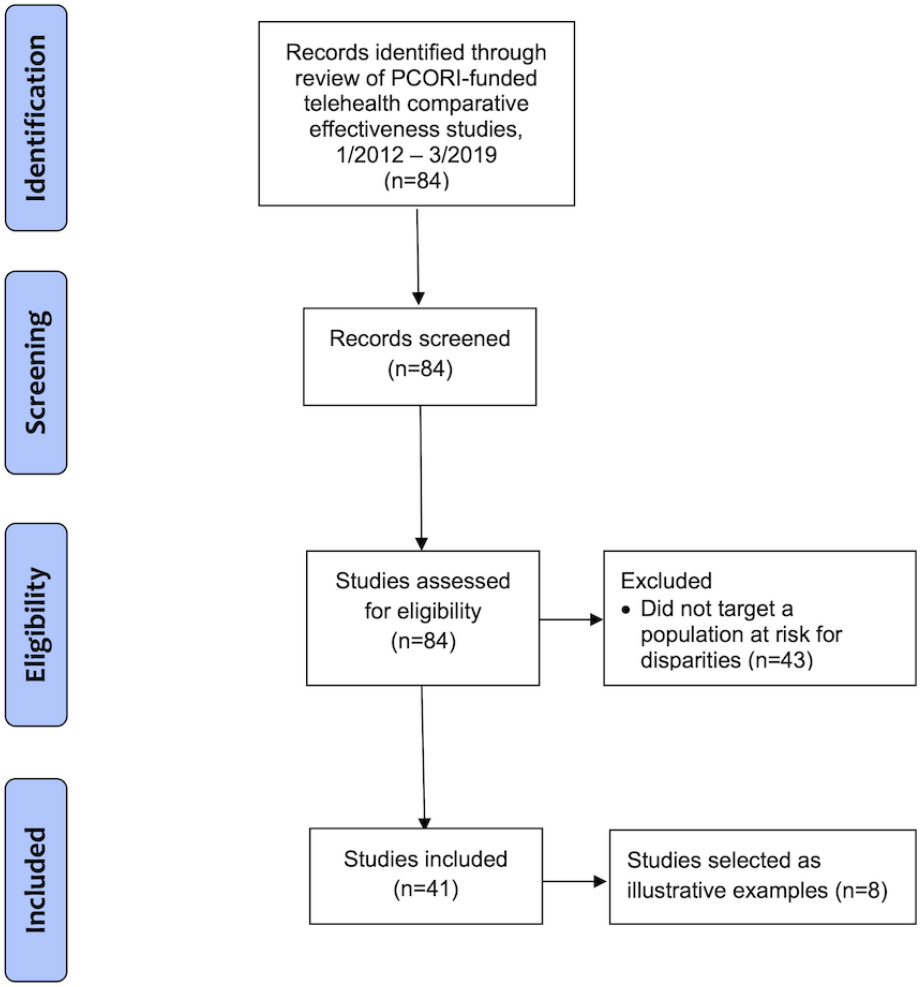
Identifying Barriers and Facilitators

We drew on these 8 illustrative examples from early PCOR experience and evidence from the literature to assess barriers and facilitators to implementing telehealth in populations at risk for disparities [25]. Key informant interviews of principal investigators were conducted for the convenience sample of 8 PCORI studies to identify common challenges in the implementation of telehealth in populations at risk for disparities. Among the 8 studies in the convenience sample, 6 principal investigators were contacted for phone interviews and 4 interviews were completed. PCORI then hosted a webinar attended by investigators from all 8 of the studies to facilitate discussion about the study challenges investigators faced and the solutions they identified and implemented. After the initial information gathering from the interviews and webinar, PCORI staff invited investigators to participate in this scoping review and share the lessons learned.

Extracting the Data

Following identification of all PCORI studies meeting the criteria for the scoping review (n=41), PCORI staff extracted descriptive data for all these studies, including telehealth purpose and modality, population, health condition, budget amount, principal investigator, and institution, using published project descriptions and study papers. PCORI staff and study investigators then extracted additional data from these same sources and interim progress reports (where available) in order to categorize documented barriers and facilitators using a published framework of identified barriers (eg, limitations) and facilitators (eg, solutions) to telehealth [25]. We then used prespecified data definitions from the Fixsen active implementation framework to further categorize barriers and facilitators to implementing telehealth within 3 major domains, including competency drivers (ie, participant selection, training, and supervision), organizational drivers (ie, decision support, administrative support, and system intervention), and leadership drivers [38,39]. The reporting process used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews (Figure 1) [40,41].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews flow diagram of the study selection process. PCORI: Patient-Centered Outcomes Research Institute.



Collating, Summarizing, and Reporting the Data

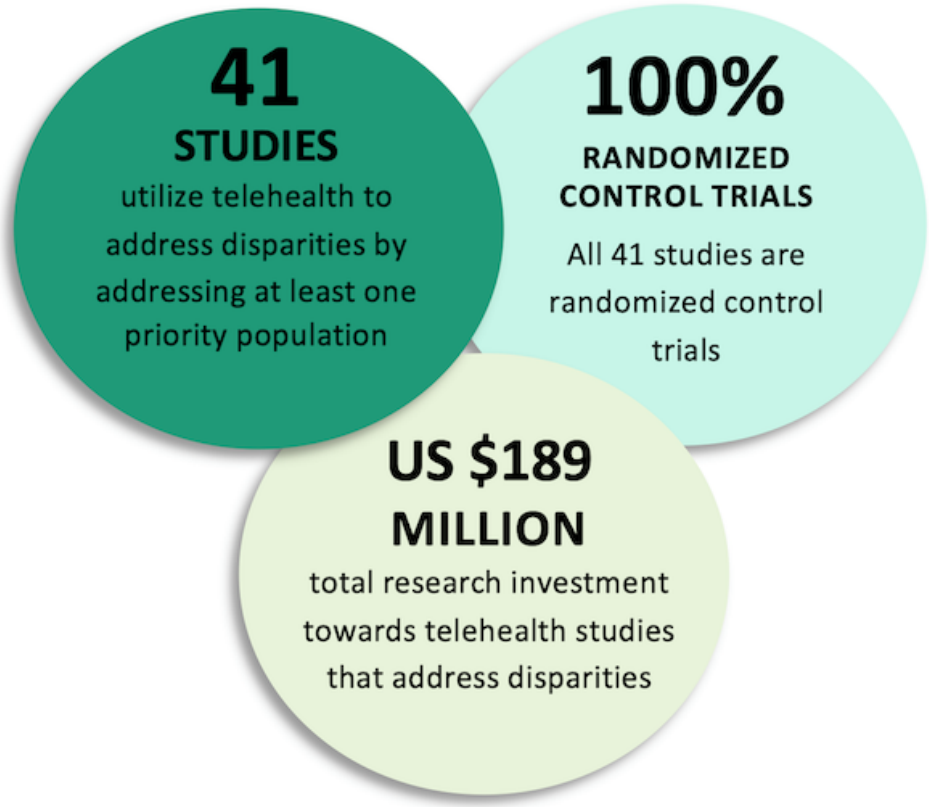
A table including the core descriptive data regarding all PCORI-funded telehealth studies addressing populations at risk for disparities was created. In addition, figures and charts were created to summarize project details about this subset of studies. An assessment of the quality of studies included in this review was not performed. A table was created summarizing the barriers and facilitators identified within each major implementation domain. Finally, major themes regarding how telehealth interventions can overcome barriers to telehealth adoption and implementation were identified through review of both the extracted data and illustrative examples using an iterative Delphi process among the 8 PCORI principal investigators participating in the study to achieve consensus. The Delphi process was facilitated by PCORI staff and a lead principal investigator for the scoping review, and was conducted through teleconference and email communication. PCORI staff and the lead principal investigator initially categorized extracted data and illustrative examples by barrier and facilitator types based on previously published categories [25]. Studies were then recategorized and category labels were revised based on feedback from participating investigators until consensus on the barrier and facilitator types was achieved.

Results

Identifying Relevant Studies

As shown in Figure 2, 41 PCORI-funded studies were identified that assessed the use of telehealth to improve outcomes for populations at risk for health or health care disparities in this analysis. Of note, all 41 studies employed a randomized controlled trial methodology, and all were pragmatic or "real-world" comparative effectiveness studies rather than efficacy studies.

Figure 2. Patient-Centered Outcomes Research Institute comparative effectiveness studies assessing telehealth solutions to address disparities.

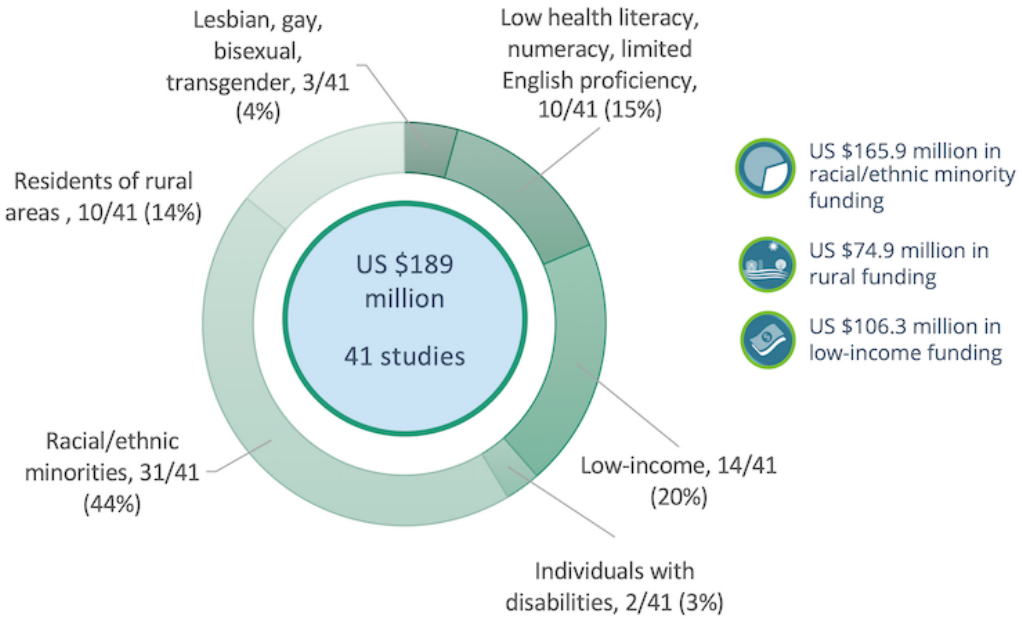


Study Characteristics

The descriptions, study goals, telehealth interventions studied, settings, number of participants, and targeted study populations for these 41 PCORI-funded studies are detailed in Multimedia Appendix 1 [42-96]. The studies ranged in size from 59 to 84,828 participants; all were conducted in ambulatory settings; and the majority (33/41, 80.5%) were conducted in primary care and/or community settings.

As shown in Figure 3, all 41 studies targeted at least one of PCORI's priority populations, including racial and ethnic minorities, people living in rural areas, people with low income or low socioeconomic status, people with low health literacy, members of the LGBT community, and people with disabilities.

Figure 3. Populations at risk of disparities targeted by Patient-Centered Outcomes Institute comparative effectiveness studies in telehealth.



PCORI-funded studies employed multiple telehealth modalities. The majority of these 41 studies used a mobile device (n=29, 71%) or were web-based (n=30, 73%). Fewer studies used real-time videoconferencing (n=15, 37%), remote patient monitoring (n=8, 20%), or store-and-forward capability (ie, asynchronous electronic transmission; n=4, 10%). It is important to note that a similar telehealth modality does not indicate similar studies. For example, studies employing mobile device interventions were diverse and included web-based interventions requiring full internet access, as well as lower-tech interventions that only employed SMS or simple text messaging. All the web-based interventions referenced below required reliable broadband internet access. Real-time videoconferencing was used in 15 studies, and 11 of these studies utilized telehealth to improve access to primary or specialty care. Health conditions involving videoconferencing were cardiovascular diseases (n=2), kidney diseases (n=2), mental/behavioral health (n=4), nutritional and metabolic disorders (n=3), allergies and immune disorders, multiple chronic conditions, neurological disorders, respiratory diseases, and trauma/injury. Remote patient monitoring (eg, blood pressure or blood glucose levels) was used in 8 studies. Remote monitoring varied from a wearable device (eg, Fitbit) to mobile app monitoring to an in-home device (eg, scale, blood pressure monitor, or video monitoring). Health conditions incorporating remote monitoring included cardiovascular diseases (n=3), mental/behavioral health (n=2), nutritional and metabolic disorders, and respiratory diseases. Store-and-forward technology was employed in 5 studies that included the following conditions: functional limitations and disabilities, mental/behavioral health, rare diseases, hearing loss, and skin diseases.

Furthermore, the purpose of the 41 PCORI telehealth interventions also varied substantially, with 8 (20%) focused on enabling health monitoring, 20 (49%) focused on improving access to specialty care, 22 (54%) focused on health education, and 27 (66%) focused on promoting chronic condition self-management. The spectrum of chronic conditions studied also varied substantially. The largest proportion of studies focused on behavioral health (n=9, 22%), and nutritional and metabolic disorders (n=6, 15%), including obesity and diabetes. In addition, PCORI has funded telehealth projects addressing disparities focused on cardiovascular conditions (n=3), chronic kidney disease (CKD; n=2), neurological conditions (n=2), reproductive and perinatal health (n=2), rare diseases (n=2), allergies and immune disorders (n=1), cancer (n=1), hearing loss and ear diseases (n=1), functional limitations and disabilities (n=1), infectious diseases (n=1), multiple chronic conditions (n=1), trauma (n=1), and skin diseases (n=1). Although only 1 study focused on patients with multiple chronic conditions, most of the studies included patients with multiple chronic conditions and some additional studies used such presence as an inclusion criterion.

Identifying Barriers and Facilitators

Barriers and facilitators of telehealth implementation in PCORI-funded studies using telehealth to address disparities are detailed in Tables 1 and 2. Barriers and facilitators are described for only 35 of the 41 studies as 6 were in the early stages of implementation and lacked sufficient data for evaluation. Review of this data revealed that 20 (57%) studies actively engaged patients to assist investigators in building patient-centered design with active patient participation in each development phase, 13 (37%) studies culturally tailored their telehealth intervention, 4 (11%) studies employed partnership with payers to expand telehealth reimbursement, and 11 (31%) studies delivered telehealth care through or with the assistance of trusted intermediaries. Of the 11 studies, 5 (14%) had clinical support team members available to support their telehealth interventions with a human interaction component.

Table 1. Barriers to telehealth adoption identified in Patient Centered Outcomes Research Institute–funded studies using telehealth to address disparities.

Barriers (ie, limitations or challenges)	Barriers identified, n	Example
Competency drivers		
Participant engagement barriers (eg, social determinants of health presenting barriers to telehealth adoption)		
Inadequate access to telecommunication technology	9	Patients with many competing social needs and low levels of social support and resources that may have prevented individuals who would otherwise be eligible and/or interested in study participation (eg, inadequate access to technology). This barrier was removed by adding "access to a phone or iPad" to the inclusion criteria.
Patient and stakeholder feedback	1	Patient and stakeholder partners were integrated into studies to provide feedback on the usability and feasibility of technology.
Staff training barriers	3	Providers and health care staff may not receive training on technology use or telehealth etiquette. Inadequate communication among patients, multiple providers, primary care providers, and specialists exacerbates the technology divide.
Inadequate program supervision/management	3	Lack of personnel structure to oversee administrators of technology platforms.
Organizational drivers		
Inadequate decision support		
Technology problems	0	Not reported.
Inadequate administrative support		
Insufficient administrative support staff to deliver or support telehealth solutions	5	Lack of support staff available to help access and utilize the different app platforms.
Health systems workflow	1	Poor integration of telehealth intervention and existing clinical care workflows.
Inadequate clinical support		
Insufficient staff to deliver telehealth solutions	5	Lack of clinical support, such as providers or lay health workers, available to assist with the delivery of telehealth platforms.
System-level barriers		
Legal (eg, inadequate safeguards to ensure patient-guided confidential protected health information sharing, state licensure laws, need for credentialing at multiple sites, and liability concerns)	2	Lack of protection for study participants, specifically those who were undocumented immigrants.
Financial (eg, limited insurance coverage for telehealth)	5	Inconsistent reimbursement coverage for providers. State-based licensure laws exist.
Leadership drivers		
Inadequate administrative leadership	0	Not reported.
Inadequate clinical leadership (eg, concerns regarding lower quality patient–physician relationship or poor physician buy-in)	0	Not reported.
Partnership	1	Partnership with national organizations to help design and implement the intervention.

Table 2. Facilitators to telehealth adoption identified in Patient Centered Outcomes Research Institute–funded studies using telehealth to address disparities.

Facilitators (ie, solutions or improvements)	Facilitators identified, n	Example
Competency drivers		
Improvements in participant engagement (eg, approaches to address social needs and facilitate telehealth adoption)		
Provision of access to telecommunication technology	4	Case managers assisted patients with addressing telehealth access barriers.
Patient-centered design (eg, telehealth solutions) actively guided by patients to meet community needs	20	Most studies engaged patients in the design of the intervention and obtained feedback from the target population.
Cultural tailoring of telehealth solutions to meet the needs of specific communities and populations	13	Culturally tailored interventions for target populations.
Improvements in staff training (eg, training in patient empowerment techniques such as motivational interviewing to facilitate telehealth adoption)	4	Technology training for providers and community health workers. Help solving workflow issues.
Organizational drivers		
Decision support solutions		
Technology solutions	0	Not reported.
Administrative solutions		
Strengthening of administrative support staff to support telehealth solution	11	Support staff to provide ongoing support and engagement for technical issues.
Process evaluation	1	Feasibility assessments to determine the fidelity of the technology.
Clinical support team solutions		
Strengthening clinical support staff to deliver or support telehealth solutions (eg, delivering telehealth through trusted intermediaries)	5	Support staff, such as nurses, providers, and lay health workers, are available to provide human interaction, which strengthens trust with technology.
System-level improvements		
Legal (eg, safeguards to ensure patient-guided confidential sharing of protected health information)	2	Reassurance for securing data using appropriate computer technology.
Financial (eg, partnering with payers to expand insurance coverage for telehealth and telehealth reimbursement)	4	Working with communities to provide coverage for community health aides to help implement technology and to better understand telehealth insurance claims.
Legislative policies	1	Experience working with state legislature/licensure to help lower barriers to telehealth.
Leadership drivers		
Improvements in administrative leadership	0	Not reported.
Improvements in clinical leadership (eg, engagement of strong physician champions to lead change)	0	Not reported.

Major themes identified across these studies included the importance of patient-centered design, cultural tailoring of telehealth solutions, delivering telehealth through trusted intermediaries, partnering with payers to expand telehealth reimbursement, and ensuring confidential sharing of private information.

Illustrative examples of these key themes are reviewed in the sections below, and key system- and policy-level changes needed to address barriers to telehealth adoption and implementation are identified.

Patient-Centered Design

Patient engagement is a core component of PCORI studies. Some studies go beyond incorporating patient input and feedback on the research and actively engage patients into the study design and overall design of the intervention.

The VIGoROUS (Video Game Rehabilitation for Outpatient Stroke) study illustrates how user-friendly consumer-driven design serves as a critical component of sustainable telehealth initiatives [70,71]. It employed a video gaming system to deliver therapeutic exercises and track adherence/progress remotely for individuals with motor disability. The gaming system was co-designed by therapists and people with motor disability. Artificial intelligence automatically adjusted the difficulty of

the exercises to match the ability of the patient, allowing for a simple user interface and less than 1-minute setup.

Similarly, the study by Williams et al comparing high-touch and high-tech care in patients with complex care needs utilized a smartphone-based remote care management platform and digital health innovations to support self-directed care management [85,97]. While the participant technology comfort level was assessed prerandomization and was expected to be variable among the study population, the study team found that participants often requested assistance in understanding basic phone functionality. Thus, the investigators worked to devise a modality flexible enough to support all levels of technical knowledge, which was easy to access, was user friendly, and addressed a broad range of skills from simple to difficult. The investigators developed a high-tech care strategy user guide with input from the stakeholder partners, patients, and providers. The need for more tailored technical support also resulted in concierge-style in-person technical visits to perform functional assessments that could not be completed over the phone, leading to a new workforce supporting additional technical initiatives across the care delivery continuum.

Cultural Tailoring of Telehealth Solutions

When developing telehealth interventions, it is particularly important to consider the need for culturally competent care, defined as care that respects the diverse characteristics and values of the patients, which can ultimately affect their overall health and health care [98]. This review of PCORI-funded studies showed that patient study engagement improved with the use of culturally competent messaging delivered via telehealth interventions.

The MODEL (Management of Diabetes in Everyday Life) study assessed the effectiveness of tailored motivational text messaging for improving diabetes self-care activities in medically underserved African American adults in the Mid-South, and illustrated the importance of cultural tailoring of telehealth solutions [44,45]. This study incorporated feedback from a patient advisory council into the message development process to more competently culturally tailor message content to the target population, an approach that has demonstrated positive impacts in other populations with diabetes [44,45]. The MODEL study's registry-based text message system employed electronic medical record and survey data, as well as ongoing message input from patients themselves to personalize and deliver culturally tailored content to patients with uncontrolled diabetes [45]. Preliminary program results indicated high satisfaction with the MODEL text message program, with greater than 90% participant retention in this group of the study.

Similarly, the Health-E You/SaludiTu study demonstrated how cultural tailoring of telehealth interventions can improve their acceptance by populations at risk of disparities [91,92]. This study aimed to assess the effectiveness of a patient-centered computer-based clinical intervention to reduce health disparities in unintended pregnancies among Latina adolescent girls in 18 school-based health centers in Los Angeles, California. Latina adolescents and local community representatives were key stakeholders in the design and implementation of this app, which may have helped the adoption and implementation of this intervention into clinic workflows at the school-based health centers. The culturally tailored tool was available in English and Spanish, and adapted language that resonated with Latina adolescents (eg, the use of the more colloquial term *condones* versus *preservativos* to refer to condoms). Preliminary findings showed that both providers and study participants reported high satisfaction with the app, as it was easily integrated into the clinic flow, and more importantly, participants showed an increase in knowledge of contraceptive use from baseline to their 6-month follow-up.

Likewise, early implementation experience from a PCORI-funded study of underserved Asian Americans having chronic hepatitis B suggested that culturally and linguistically competent interventions improve participant retention [72]. This study employed patient-designed culturally tailored text messaging along with patient navigator–led educational sessions to improve adherence to medication therapy for hepatitis B. Patient partners and community stakeholder groups provided extensive input on the appropriateness and framing of the messages. These messages were delivered in Chinese, Vietnamese, or Korean. Early implementation findings showed evidence that this culturally and linguistically appropriate intervention was effective in improving follow-up in patients living with hepatitis B.

Delivering Telehealth Through Trusted Intermediaries

Many of the studied PCORI telehealth projects successfully employed trusted intermediaries to deliver telehealth solutions and help support people at high risk for health disparities in crossing the digital divide. These projects provided confirmatory evidence to extensive literature suggesting that lay health worker and paraprofessional intermediaries can help build trust among patients and providers in new telehealth modalities. New technologies, including telehealth, that enable the delivery of interventions and diagnostics remotely are emerging, but patients commonly struggle to use these technologies. Patients struggle with both simple technical issues, such as neglecting to ensure that an electronic system is plugged in, and more complex barriers related to social determinants of health (eg, poverty, unpredictable work schedules, unreliable transportation, and lack of internet access) when using advanced technologies such as secure videoconferencing.

PCORI telehealth projects consistently demonstrate high patient acceptance of telehealth modalities that connect them to trusted providers, but even with careful attention to patient-centered design, investigators commonly found availability of in-person technological support to be essential. For example, when 1 participant in the aforementioned VIGoROUS video gaming intervention reported that her gaming system would not turn on, a support call revealed that she had not plugged it into an electrical outlet [70,71]. This early experience suggested that widescale implementation of telehealth technologies requires funding of responsive technical support infrastructure to avoid technology abandonment by patients in greatest need.

Similarly, a study of Native Americans addressing gaps in self-management through home-based CKD care found that continuous care coordination using trained and trusted tribal community health representatives (CHRs) was essential to

effective program implementation [86-88]. This program used videoconferencing with patients via internet-connected portable tablets using Verizon wireless Jetpack mobile hotspots. The program worked with other local health and wellness programs (eg, Indian Health Services and other programs operating within tribal communities) and conducted community-wide health fairs. Telehealth adoption was also encouraged through distribution of culturally sensitive newsletters and brochures to all these health programs and individual community members. The secondary data analysis of the study demonstrated that behavioral and lifestyle educational reinforcement through motivational text messaging and alternate weekly home visits by the CHRs with quarterly group sessions was an effective means of providing care to patients with type 2 diabetes mellitus and CKD who may otherwise avoid diagnosis and treatment due to stigmatization. The CHR-led home-based care model provided the additional care necessary to bolster patient levels of disease-specific knowledge, self-efficacy, and diabetes and CKD self-management, enabling patients to more effectively carry out the recommendations that they received during the home visits, as compared to patients who received clinic-based usual care [99].

A study addressing childhood hearing loss in an Alaska Native population similarly found that using trusted lay health workers based in the communities served was essential to telehealth program deployment [100-103]. In the remote rural communities served by this program, health care is provided by community health aides/practitioners who are commonly from the community they serve and are selected by their community to receive training. This care is supported by specialist triage using the statewide telemedicine network. As a result, ear and hearing problems in isolated communities are routinely managed remotely by audiologists and surgeons through telemedicine, including provision of local care (such as for ear infections), preoperative planning, and postoperative follow-up [102,103]. This requires community health aides/practitioners performing telemedicine consults to be proficient with telemedicine equipment, including basic testing, such as tympanometry, to assess eardrum movement and digital otoscopy for photos of the eardrum. Training with telemedicine equipment is built into each session of the standardized community health aide/practitioner program and includes continued on-the-job training and direct observation of practical skills. Expanding the use of telemedicine equipment for prevention, such as follow-up for referred school hearing screenings, utilizes community health aide/practitioner knowledge and training on the telemedicine infrastructure already in place. However, Emmett et al also provided additional training on the expedited telemedicine consult form to facilitate the application of modified workflows for school hearing screening referrals [68,69].

Partnering With Payers to Expand Telehealth Reimbursement

Reimbursement for telehealth encounters is increasing. However, lack of adequate reimbursement continues to present a major challenge to telehealth adoption, implementation, and sustainability. Reimbursement barriers to telehealth were acknowledged by nearly all the PCORI-funded study investigators who participated in this study. Despite recent expansions in telehealth reimbursement during the COVID-19 pandemic, variable state and federal reimbursement policies and low levels of reimbursement for services delivered via telehealth continue to challenge telehealth implementation efforts. In response to these challenges, many studies found it essential to partner with payers to expand their support and/or reimbursement for telehealth services.

For example, the TEAM UP (Treatment Efforts Addressing Child Weight Management by Unifying Patients, Parents, and Providers) study assembled a Payer Advisory Board that explored how military medical sites use telehealth for services such as teleradiology and telebehavioral health [96]. The Board noted how telehealth could facilitate access to qualified and trained health care professionals in rural areas, as well as foster treatment engagement and reduce program attrition by increasing flexibility for patients with challenging schedules. The Payer Advisory Board stressed the viability of telehealth and the need for these types of visits to be considered billable encounters to help overcome the challenges facing populations with barriers to care.

In the rural Alaska study addressing childhood hearing loss, the research team developed a streamlined telemedicine consultation process in order to make it financially sustainable in the school setting [68,69,104]. Translating a clinically oriented technology to a preventive service raises inherent logistical challenges [69]. The team shortened the telemedicine consult time from 60-90 minutes to 10-15 minutes, making it feasible to perform 10 to 15 telemedicine referrals for school hearing screening in a single day. This process took several iterations, with feedback sought from multiple stakeholders including community health aides/practitioners. Maintaining the integrity of billing requirements, while shortening the telemedicine consultation process, allowed the preventive service to remain reimbursable and financially sustainable after the conclusion of the randomized trial.

Similarly, in the Mid-South MODEL study, investigators worked with regional health systems to provide long-term support for a regional registry-based tailored text message system that can assist both large health systems and small independent primary care practices in sending regular motivational messages to interested patients [44,45].

Ensuring Confidential Sharing of Private Information

Early implementation experience from numerous PCORI telehealth studies demonstrates the importance of ensuring confidential sharing of private information in all telehealth interventions considered. The application of telehealth often involves paperless collection and transmission of data at each touch point with participants and with communication between staff at the regional and local levels. This requires careful attention to compliance with local, regional, and federal legal requirements for protection of privacy.

For example, the TEAM UP study of alternative childhood obesity treatments for Medicaid-insured individuals employed a secure Health Insurance Portability and Accountability Act (HIPAA)-compliant videoconferencing system to enable remote

communication with children and families that protected privacy [96]. Working with a low-income population presents an inherent set of barriers, including unpredictable schedules and unreliable transportation, both of which could negatively impact the fidelity of family-based behavioral treatment delivery. In anticipation of these potential barriers, the study team contracted with a fully HIPAA-compliant videoconferencing system to allow for telehealth delivery of family-based behavioral treatment sessions as needed. Secure video was found to be an ideal technology solution since it is accessible on virtually all devices, including cell phones, and provides necessary flexibility to support adherence with demanding treatment schedules. Use of secure HIPAA-compliant videoconferencing ensured the confidentiality and security of participants' protected health information in compliance with part 11 of Title 21 of the Code of Federal Regulations pertaining to Electronic Records and Electronic Signatures (21 CFR Part 11) [96]. Likewise, the study employed an electronic consent process that allows for the use of alternative consent forms for each participating site.

The VIGoROUS study also had to address challenges related to information security [70]. First, it was necessary to find a HIPAA-compliant videoconferencing platform for secure video chat between participants and providers. While there are numerous platforms available, some of which are free, the lead site's hospital system was relatively unfamiliar with telehealth and thus required new technology platforms to undergo an extensive information security vetting process (eg, conducting risk assessments and obtaining administrative approvals), a process that ultimately spanned 3 years. To proceed with the study, investigators were able to identify and pivot to a user-friendly videoconferencing platform (Bluejeans) that had already been vetted by the university health system. The second challenge was how to make large amounts of remote-monitoring data (adherence data, quality of movement data generated continuously during rehabilitation, and behavioral assessment data) accessible to a clinician providing care remotely. This was addressed by streaming deidentified data to a secure cloud server. The data were stored under a unique machine identifier that could only be obtained through physically accessing the rehabilitation gaming system (located within participants' homes). During video consultations with a therapist, the participant provided the therapist with the machine identifier to enable the therapist to track his/her progress [70].

Discussion

Principal Findings

This scoping review of early PCOR evidence suggests that the most effective health system- and provider-level telehealth implementation solutions to address disparities employ patient-centered and culturally tailored telehealth solutions. We found that the development of the most effective telehealth solutions was actively guided by patients themselves in order to meet the needs of specific communities and populations. Early PCOR evidence demonstrates that the best practices in telehealth implementation include delivering telehealth through trusted intermediaries, close partnership with payers to facilitate reimbursement and sustainability, and safeguards to ensure patient-guided confidential sharing of personal health information.

The COVID-19 pandemic has given new urgency to these questions regarding telehealth adoption among people at the highest risk for disparities, including those living in medically underserved, low-income, or rural areas; those from racial and ethnic minorities; LGBT persons; and those with limited English proficiency or disabilities. Telehealth modalities have great potential to help overcome these geographic, socioeconomic, cultural, and language barriers and to give populations at risk of disparities enhanced access to essential health services. Although many PCORI-funded studies had to put their research on hold and transfer their institution's focus on the needs of patients experiencing COVID-19 [105], the pandemic generally helped to accelerate telehealth adoption for vulnerable populations [26,28]. However, since early evidence indicates that differences in internet and telehealth access may worsen disparities in chronic disease and COVID-19 outcomes [106], it is clear that telehealth solutions must deliberately target and prioritize populations at the highest risk for disparities.

Toward that end, this review places strong emphasis on the importance of deploying patient-centered and culturally tailored telehealth solutions that address the social determinants of health faced by populations at risk for disparities [19-21]. Even though low-income, minority, and rural populations face heightened barriers to effective self-care related to the social determinants of health, most telehealth behavioral interventions for chronic conditions are tailored for individuals with higher incomes. Thus, low-income, minority, and rural groups need specific culturally tailored solutions that specifically address the social determinants of health they face [69,104]. Preliminary evidence from PCORI-funded studies consistently demonstrates that intensive personalization and cultural tailoring of intervention components [70] can help enable vulnerable patients to address critical social determinants of health [71]. Effective telehealth interventions need to address the key social determinants of health at the root of entrenched health behaviors in systematic ways [72]. Most of the reviewed PCOR initiatives employed component interventions that were extensively culturally tailored during the initial program planning phases and on an ongoing basis to ensure that they were culturally congruent and appropriate based on the subjective culture (ie, norms and attitudes), behavioral preferences, and cultural values and expectations of the population served [71,73]. Thus, although cultural tailoring alone is not sufficient, the early PCOR literature indicates that it is an essential component of effective telehealth solutions to address health disparities.

Furthermore, this review revealed that telehealth solutions can take advantage of and expand on existing technological capacities accessible to low-income and other populations at risk for health disparities, and can often be deployed at low cost. The existing PCORI-funded telehealth research suggests that low-cost strategies that employ existing telehealth capacities using standard mobile phones and smartphones may be particularly effective in reaching patients where they are and engaging them in needed care. Although more than 90% of Americans overall carry cell phones and 80% have smartphones [107], among people with household incomes less than US

$30,000, only 71% own a smartphone, 54% own some type of computer, and 56% have home broadband [106]. As demonstrated by several of our case studies above, early evidence indicates that text messaging and other culturally tailored mobile health (mHealth) interventions can provide effective low-cost approaches for engaging patients in self-care [108-112]. Some systematic reviews have identified text messaging as among the most effective low-cost technological strategies for engaging patients in behavior change and have highlighted the importance of such technology-supported behavioral interventions [113-115]. Thus, health systems seeking to target telehealth solutions to address disparities should actively seek to employ existing technologies and devices like traditional cell phones to which populations at risk for disparities already have access.

This review also emphasizes the importance of telehealth solutions that work through existing trusted care networks and providers, and that employ trusted lay health worker and paraprofessional intermediaries to introduce and support the use of these technologies by people from the community. Many of the most effective PCORI telehealth initiatives considered in this paper focused their interventions around engaging, training, and deploying trusted intermediaries from the community and/or the population at risk for disparities to be served. This evidence also suggests that personalized human interaction is essential. For example, text messaging interactions were found to be most effective when delivered in real-time, tailored to participant interests and needs, and originated from a trusted known source. The PCORI-funded telehealth projects reviewed suggested that face-to-face videoconferencing and telephone interventions with trusted caregivers are particularly effective.

Our review further suggests that substantive reimbursement and regulatory changes are immediately needed to enable low-cost and efficient telehealth solutions. We found that the best practices in telehealth implementation include close partnership with payers to facilitate reimbursement and sustainability. However, ultimately, experience from the existing portfolio of PCORI telehealth projects indicates that permanent changes are needed in national reimbursement policies and regulations to facilitate broader telehealth adoption and implementation. Policies and reimbursement for telehealth were rapidly put into place to allow health care to continue during the severe restrictions early in the COVID-19 public health emergency after March 2020, in order to allow care to continue during the COVID-19 pandemic. However, these changes clearly need to be extended. Needed regulatory changes include expanded reimbursement as well as relaxations of regulations around where the patient or practitioner is located, licensure, and reciprocity across states. For example, since populations at risk for disparities generally have less access to the internet and technology, regulatory solutions need to facilitate and enable the use of existing lower-cost technologies, such as telephone visits using traditional cell phones, rather than requiring video teleconferencing for medical billing. As the pandemic eases, it will be essential to carefully think about how to preserve the progress that was made in expanding telehealth in ways that improve patient care at the federal (eg, Medicare), state, and individual insurance company levels [116-118].

Strengths and Limitations

The existing PCORI telehealth research targeting populations at risk for health disparities is of very high quality and has high applicability to real-world clinical and community settings. Of note, 100% of the PCORI-funded studies in this category employed a randomized controlled trial design and 100% were pragmatic rather than being conducted in highly controlled university settings among highly selected patients. This is a major strength of both the research reviewed in this scoping review and the scoping review itself. Furthermore, the studies included in this scoping review are also notable for their generalizability, given their pragmatic real-world approach. Rather than excluding patients with multiple chronic conditions as is often done in randomized clinical trials, many of the PCORI-funded studies specifically included patients with multiple morbidities given their representativeness of patients most commonly seen in real-world clinical settings.

This review is limited by the small numbers of studies that were considered, and its main findings are subject to numerous potential biases. First, since PCORI staff chose studies for an in-depth review based on their perception regarding which study teams overcame notable obstacles to telehealth, selection bias could have occurred. Our expectation is that this bias would most likely lead to the overemphasis of facilitators and solutions as opposed to barriers and challenges. Second, because PCORI staff and investigators comprised the study team, it is possible that there was some bias toward reporting of positive study findings. However, the study team and methods actively sought to address this potential limitation by identifying and categorizing barriers first prior to the discussion of solutions. In addition, some of the telehealth services provided in these studies may be difficult to consistently replicate outside of structured evaluations.

The review is also limited because final study results are not available for many of the included studies, and even fewer have detailed analyses of implementation factors and lessons learned available in the peer-reviewed literature. However, the review was facilitated by the detailed project knowledge of PCORI staff and the investigators themselves.

Future Directions

Over the past 5 years, PCORI has supported many comparative effectiveness studies of adopting virtual care solutions to manage health outcomes across a wide range of populations at risk for disparities. Through these research initiatives, health care delivery systems are learning to leverage low-cost telehealth solutions to promote better health outcomes by increasing access to care, making care more effective, and continuously engaging patients. PCORI-funded comparative effectiveness research is demonstrating how we can reduce health disparities across the spectrum of disease conditions and populations through low-cost telehealth solutions. This research is identifying key barriers and limitations to implementing telehealth in vulnerable populations and ways to overcome these barriers to help improve health and health care outcomes.

In this new era of the COVID-19 pandemic, there has been rapid growth in the use of technology across all industries, but it is proving to be particularly important in the areas of health and health care. Future efforts should focus on fostering collaborations between researchers, payers, and administrators to facilitate widespread adoption of new treatment models that have already demonstrated success through definitive clinical trials. Our review indicates that key system- and policy-level changes are desperately needed to address barriers to telehealth reimbursement. Moreover, further pragmatic research is needed to demonstrate the best approaches for disseminating and scaling these early telehealth findings in broader community practice to promote population health.

Conclusions

This scoping review gives strong guidance to health systems seeking to target evidence-based telehealth modalities for health disparities. Results from the studies demonstrate that systems can do so in innovative ways, and in fact, some of the studies identify methods for systems to effectively reach underserved populations in their own communities. These case studies also highlight the critical importance of both supportive infrastructure, and regulatory and reimbursement policies to facilitate telehealth modalities and make them sustainable.

However, despite growing evidence that telehealth can deliver care to people at risk for disparities, many major obstacles exist for its implementation. Early PCOR evidence demonstrates that the most effective health system- and provider-level solutions for at-risk populations use patient-centered and culturally tailored telehealth solutions whose development is actively guided by the patients themselves.

Acknowledgments

The authors gratefully acknowledge the many Patient-Centered Outcomes Research Institute (PCORI) study participant patient advisors who contributed their substantial expertise on patient experience to this scoping review. Without their participation and insights, this study would not have been possible. We are also very thankful for the contributions of the numerous PCORI-funded investigators and teams whose studies are cited in this scoping review. Lastly, we specifically acknowledge the editorial contributions of James Schuster, MD, at the University of Pittsburgh Medical Center, who also serves on the Board of Governors of the PCORI.

Conflicts of Interest

All listed authors are either investigators associated with Patient-Centered Outcomes Research Institute (PCORI)-funded studies in its specific telehealth portfolio, PCORI employees, or PCORI contractors/consultants. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, or its Board of Governors or Methodology Committee.

Multimedia Appendix 1
Patient-Centered Outcomes Research Institute–funded studies assessing the use of telehealth to improve outcomes for populations at risk for health or health care disparities.
[DOCX File , 38 KB - jmir_v23i12e28503_app1.docx]

References

1. James CV, Moonesinghe R, Wilson-Frederick SM, Hall JE, Penman-Aguilar A, Bouye K. Racial/Ethnic Health Disparities Among Rural Adults - United States, 2012-2015. MMWR Surveill Summ 2017 Nov 17;66(23):1-9 [FREE Full text] [doi: 10.15585/mmwr.ss6623a1] [Medline: 29145359]
2. Bach PB. Book Review Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care Edited by Brian D. Smedley, Adrienne Y. Stith, and Alan R. Nelson. 764 pp. Washington, D.C., National Academies Press, 2003. $79.95. 0-309-08532-2. N Engl J Med 2003 Sep 25;349(13):1296-1297. [doi: 10.17226/12875] [Medline: 25032386]
3. Semega J, Kollar M, Shrider E, Creamer J. Income and Poverty in the United States: 2019. United States Census Bureau. URL: https://www.census.gov/library/publications/2020/demo/p60-270.html [accessed 2020-09-24]
4. Preventing Chronic Diseases and Promoting Health in Rural Communities. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/chronicdisease/resources/publications/factsheets/rural-health.htm [accessed 2020-12-15]
5. Bolin JN, Bellamy GR, Ferdinand AO, Vuong AM, Kash BA, Schulze A, et al. Rural Healthy People 2020: New Decade, Same Challenges. J Rural Health 2015;31(3):326-333. [doi: 10.1111/jrh.12116] [Medline: 25953431]
6. Blewett LA, Johnson PJ, Lee B, Scal PB. When a usual source of care and usual provider matter: adult prevention and screening services. J Gen Intern Med 2008 Sep 28;23(9):1354-1360 [FREE Full text] [doi: 10.1007/s11606-008-0659-0] [Medline: 18506542]
7. Kim MY, Kim JH, Choi I, Hwang IH, Kim SY. Effects of having usual source of care on preventive services and chronic disease control: a systematic review. Korean J Fam Med 2012 Nov;33(6):336-345 [FREE Full text] [doi: 10.4082/kjfm.2012.33.6.336] [Medline: 23267419]

8. Keppel KG. Ten largest racial and ethnic health disparities in the United States based on Healthy People 2010 Objectives. Am J Epidemiol 2007 Jul 01;166(1):97-103. [doi: 10.1093/aje/kwm044] [Medline: 17463050]
9. Conway BN, Han X, Munro HM, Gross AL, Shu X, Hargreaves MK, et al. The obesity epidemic and rising diabetes incidence in a low-income racially diverse southern US cohort. PLoS One 2018;13(1):e0190993 [FREE Full text] [doi: 10.1371/journal.pone.0190993] [Medline: 29324894]
10. Ogden CL, Carroll MD, Fakhouri TH, Hales CM, Fryar CD, Li X, et al. Prevalence of Obesity Among Youths by Household Income and Education Level of Head of Household - United States 2011-2014. MMWR Morb Mortal Wkly Rep 2018 Feb 16;67(6):186-189 [FREE Full text] [doi: 10.15585/mmwr.mm6706a3] [Medline: 29447142]
11. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of Obesity Among Adults and Youth: United States, 2015-2016. NCHS Data Brief 2017 Oct(288):1-8 [FREE Full text] [Medline: 29155689]
12. Dewees S, Marks B. Twice Invisible: Understanding Rural Native America. First Nations Development Institute Research Note. 2017. URL: https://www.usetinc.org/wp-content/uploads/bvenuti/WWS/2017/May%202017/May%208/Twice%20Invisible%20-%20Research%20Note.pdf [accessed 2021-10-10]
13. Marrone S. Understanding barriers to health care: a review of disparities in health care services among indigenous populations. Int J Circumpolar Health 2007 Jun 15;66(3):188-198. [doi: 10.3402/ijch.v66i3.18254] [Medline: 17655060]
14. Disparities. Indian Health Services. URL: https://www.ihs.gov/newsroom/factsheets/disparities/ [accessed 2021-02-22]
15. Hafeez H, Zeshan M, Tahir MA, Jahan N, Naveed S. Health Care Disparities Among Lesbian, Gay, Bisexual, and Transgender Youth: A Literature Review. Cureus 2017 Apr 20;9(4):e1184 [FREE Full text] [doi: 10.7759/cureus.1184] [Medline: 28638747]
16. Hickey KT, Masterson Creber RM, Reading M, Sciacca RR, Riga TC, Frulla AP, et al. Low health literacy: Implications for managing cardiac patients in practice. Nurse Pract 2018 Aug;43(8):49-55 [FREE Full text] [doi: 10.1097/01.NPR.0000541468.54290.49] [Medline: 30028773]
17. Proctor K, Wilson-Frederick SM, Haffer SC. The Limited English Proficient Population: Describing Medicare, Medicaid, and Dual Beneficiaries. Health Equity 2018 Jun;2(1):82-89 [FREE Full text] [doi: 10.1089/heq.2017.0036] [Medline: 30283853]
18. Sentell T, Braun KL. Low health literacy, limited English proficiency, and health status in Asians, Latinos, and other racial/ethnic groups in California. J Health Commun 2012;17 Suppl 3:82-99 [FREE Full text] [doi: 10.1080/10810730.2012.712621] [Medline: 23030563]
19. Butkus R, Rapp K, Cooney TG, Engel LS. Envisioning a Better U.S. Health Care System for All: Reducing Barriers to Care and Addressing Social Determinants of Health. Ann Intern Med 2020 Jan 21;172(2_Supplement):S50-S59. [doi: 10.7326/m19-2410]
20. Bierman AS, Dunn JR. Swimming upstream. Access, health outcomes, and the social determinants of health. J Gen Intern Med 2006 Jan;21(1):99-100 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00317.x] [Medline: 16423133]
21. Hill-Briggs F, Adler NE, Berkowitz SA, Chin MH, Gary-Webb TL, Navas-Acien A, et al. Social Determinants of Health and Diabetes: A Scientific Review. Diabetes Care 2020 Nov 02;44(1):258-279. [doi: 10.2337/dci20-0053] [Medline: 33139407]
22. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
23. Syed ST, Sharp LK, Kim Y, Jentleson A, Lora CM, Touchette DR, et al. Relationship Between Medication Adherence and Distance to Dispensing Pharmacies and Prescribers Among an Urban Medicaid Population with Diabetes Mellitus. Pharmacotherapy 2016 Jun 28;36(6):590-597 [FREE Full text] [doi: 10.1002/phar.1757] [Medline: 27087250]
24. Huilgol Y, Joshi A, Carr B, Hollander J. Giving Urban Health Care Access Issues The Attention They Deserve In Telemedicine Reimbursement Policies. Health Affairs Blog. 2017. URL: https://www.healthaffairs.org/do/10.1377/hblog20171022.713615/full/ [accessed 2021-11-12]
25. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161 [FREE Full text] [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
26. Demeke HB, Merali S, Marks S, Pao LZ, Romero L, Sandhu P, et al. Trends in Use of Telehealth Among Health Centers During the COVID-19 Pandemic - United States, June 26-November 6, 2020. MMWR Morb Mortal Wkly Rep 2021 Feb 19;70(7):240-244 [FREE Full text] [doi: 10.15585/mmwr.mm7007a3] [Medline: 33600385]
27. Rodriguez-Guzman E, Aktaruzzaman U, Baker A, DiBitetto K, Lima A, Lindros J, et al. Engaging key stakeholders in a pragmatic, family-centered childhood obesity study in primary care practices. Pediatrics 2020;146(Issue 1_MeetingAbstract):84-85. [doi: 10.1542/peds.146.1MA1.84]
28. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan 14;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
29. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]

30. Dellifraine JL, Dansky KH. Home-based telehealth: a review and meta-analysis. J Telemed Telecare 2008;14(2):62-66. [doi: 10.1258/jtt.2007.070709] [Medline: 18348749]
31. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
32. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015 Jan;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]
33. Vadheim LM, Patch K, Brokaw SM, Carpenedo D, Butcher MK, Helgerson SD, et al. Telehealth delivery of the diabetes prevention program to rural communities. Transl Behav Med 2017 Jun;7(2):286-291 [FREE Full text] [doi: 10.1007/s13142-017-0496-y] [Medline: 28417426]
34. Xu T, Pujara S, Sutton S, Rhee M. Telemedicine in the Management of Type 1 Diabetes. Prev Chronic Dis 2018 Jan 25;15:E13 [FREE Full text] [doi: 10.5888/pcd15.170168] [Medline: 29369757]
35. Bhatt J, Bathija P. Ensuring Access to Quality Health Care in Vulnerable Communities. Acad Med 2018 Sep;93(9):1271-1275 [FREE Full text] [doi: 10.1097/ACM.0000000000002254] [Medline: 29697433]
36. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
37. What and Who We Fund. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/funding-opportunities/what-who-we-fund [accessed 2020-09-23]
38. Fixsen D, Blase K, Naoom S, Duda M. Implementation Drivers: Assessing Best Practices. Chapel Hill, NC: National Implementation Science Network; 2013.
39. Fixsen D, Naoom S, Blase K, Friedman R, Wallace F. Implementation Research: A Synthesis of the Literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network; 2005.
40. 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]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
42. Armstrong AW, Chambers CJ, Maverakis E, Cheng MY, Dunnick CA, Chren M, et al. Effectiveness of Online vs In-Person Care for Adults With Psoriasis: A Randomized Clinical Trial. JAMA Netw Open 2018 Oct 05;1(6):e183062 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.3062] [Medline: 30646223]
43. Ford AR, Gibbons CM, Torres J, Kornmehl HA, Singh S, Young PM, et al. Access to Dermatological Care with an Innovative Online Model for Psoriasis Management: Results from a Randomized Controlled Trial. Telemed J E Health 2019 Jul;25(7):619-627 [FREE Full text] [doi: 10.1089/tmj.2018.0160] [Medline: 30222518]
44. Bailey JE, Surbhi S, Gatwood J, Butterworth S, Coday M, Shuvo SA, et al. The management of diabetes in everyday life study: Design and methods for a pragmatic randomized controlled trial comparing the effectiveness of text messaging versus health coaching. Contemp Clin Trials 2020 Sep;96:106080. [doi: 10.1016/j.cct.2020.106080] [Medline: 32653539]
45. 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]
46. Blalock SJ, Solow EB, Reyna VF, Keebler M, Carpenter D, Hunt C, et al. Enhancing Patient Understanding of Medication Risks and Benefits. Arthritis Care Res (Hoboken) 2020 Aug 16:1-26. [doi: 10.1002/acr.24421] [Medline: 32799397]
47. Remote Cognitive Behavior Therapy for Major Depression in Primary Care. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/remote-cognitive-behavior-therapy-major-depression-primary-care [accessed 2020-10-25]
48. Comparing Two Ways to Improve Depression among African-American and/or Latino LGBTQ Adults. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2016/comparing-two-ways-improve-depression-among-african-american-andor-latino [accessed 2020-10-25]
49. Coker TR, Porras-Javier L, Zhang L, Soares N, Park C, Patel A, et al. A Telehealth-Enhanced Referral Process in Pediatric Primary Care: A Cluster Randomized Trial. Pediatrics 2019 Mar 15;143(3):e20182738. [doi: 10.1542/peds.2018-2738] [Medline: 30770523]
50. Contreras S, Porras-Javier L, Zima BT, Soares N, Park C, Patel A, et al. Development of a Telehealth-Coordinated Intervention to Improve Access to Community-Based Mental Health. Ethn Dis 2018 Sep 06;28(Supp):457-466. [doi: 10.18865/ed.28.s2.457]
51. Comparing Ways to Help Sexual and Gender Minority Male Survivors of Sexual Abuse Reduce Distress and Engage in Treatment. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-ways-help-sexual-and-gender-minority-male-survivors-sexual-abuse [accessed 2020-12-15]

52. Comparing Two Methods to Deliver a Health Education Program for Pregnant Women Treated for Opioid Use Disorder in Rural Kentucky -- The PATH Home Trial. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-two-methods-deliver-health-education-program-pregnant-women-treated [accessed 2020-10-25]
53. Dehlendorf C, Fitzpatrick J, Fox E, Holt K, Vittinghoff E, Reed R, et al. Cluster randomized trial of a patient-centered contraceptive decision support tool, My Birth Control. Am J Obstet Gynecol 2019 Jun;220(6):565.e1-565.e12. [doi: 10.1016/j.ajog.2019.02.015] [Medline: 30763545]
54. Comparing Three Ways to Obtain Hearing Aids for People with Age-Related Hearing Loss. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2019/comparing-three-ways-obtain-hearing-aids-people-age-related-hearing-loss [accessed 2020-12-15]
55. Beck CA, Beran DB, Biglan KM, Boyd CM, Dorsey ER, Schmidt PN, Connect.Parkinson Investigators. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology 2017 Sep 12;89(11):1152-1161 [FREE Full text] [doi: 10.1212/WNL.0000000000004357] [Medline: 28814455]
56. Dorsey ER, Achey MA, Beck CA, Beran DB, Biglan KM, Boyd CM, et al. National Randomized Controlled Trial of Virtual House Calls for People with Parkinson's Disease: Interest and Barriers. Telemed J E Health 2016 Jul;22(7):590-598 [FREE Full text] [doi: 10.1089/tmj.2015.0191] [Medline: 26886406]
57. Achey MA, Beck CA, Beran DB, Boyd CM, Schmidt PN, Willis AW, et al. Virtual house calls for Parkinson disease (Connect.Parkinson): study protocol for a randomized, controlled trial. Trials 2014 Nov 27;15:465 [FREE Full text] [doi: 10.1186/1745-6215-15-465] [Medline: 25431346]
58. Carroll JK, Tobin JN, Luque A, Farah S, Sanders M, Cassells A, et al. "Get Ready and Empowered About Treatment" (GREAT) Study: a Pragmatic Randomized Controlled Trial of Activation in Persons Living with HIV. J Gen Intern Med 2019 Sep 25;34(9):1782-1789 [FREE Full text] [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
59. Fiscella K, Boyd M, Brown J, Carroll J, Cassells A, Corales R, et al. Activation of persons living with HIV for treatment, the great study. BMC Public Health 2015 Oct 16;15(1):1056 [FREE Full text] [doi: 10.1186/s12889-015-2382-1] [Medline: 26474979]
60. Fortney JC, Heagerty PJ, Bauer AM, Cerimele JM, Kaysen D, Pfeiffer PN, et al. Study to promote innovation in rural integrated telepsychiatry (SPIRIT): Rationale and design of a randomized comparative effectiveness trial of managing complex psychiatric disorders in rural primary care clinics. Contemp Clin Trials 2020 Mar;90:105873. [doi: 10.1016/j.cct.2019.105873] [Medline: 31678410]
61. Cerimele JM, LePoire E, Fortney JC, Hawrilenko M, Unützer J, Bauer AM. Bipolar disorder and PTSD screening and telepsychiatry diagnoses in primary care. Gen Hosp Psychiatry 2020 Jul;65:28-32. [doi: 10.1016/j.genhosppsych.2020.05.006] [Medline: 32447194]
62. Comparing Online and In-Person Cognitive Behavioral Therapy in Youth with Anxiety Disorders -- The Kids FACE FEARS Study. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-online-and-person-cognitive-behavioral-therapy-youth-anxiety [accessed 2020-10-30]
63. Pekmezaris R, Kozikowski A, Pascarelli B, Wolf-Klein G, Boye-Codjoe E, Jacome S, et al. A Telehealth-Delivered Pulmonary Rehabilitation Intervention in Underserved Hispanic and African American Patients With Chronic Obstructive Pulmonary Disease: A Community-Based Participatory Research Approach. JMIR Form Res 2020 Jan 31;4(1):e13197 [FREE Full text] [doi: 10.2196/13197] [Medline: 32012039]
64. Comparing telehealth to outpatient rehabilitation for Hispanic and African-American patients with chronic obstructive pulmonary disease (COPD). Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2016/testing-whether-referral-telehealth-program-helps-african-american-and [accessed 2020-10-29]
65. Comparing Ways to Promote Quitting Smoking for People Referred for Lung Cancer Screenings. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-ways-promote-quitting-smoking-people-referred-lung-cancer-screenings [accessed 2020-12-15]
66. Comparing Two Ways to Help Patients with Sickle Cell Disease Manage Pain -- CaRISMA. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2019/comparing-two-ways-help-patients-sickle-cell-disease-manage-pain [accessed 2020-12-15]
67. Comparing Three Methods to Help Patients Manage Type 2 Diabetes. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2013/comparing-three-methods-help-patients-manage-type-2-diabetes [accessed 2020-10-30]
68. Emmett SD, Robler SK, Wang N, Labrique A, Gallo JJ, Hofstetter P. Hearing Norton Sound: a community randomised trial protocol to address childhood hearing loss in rural Alaska. BMJ Open 2019 Jan 15;9(1):e023078 [FREE Full text] [doi: 10.1136/bmjopen-2018-023078] [Medline: 30782695]
69. Emmett SD, Robler SK, Gallo JJ, Wang N, Labrique A, Hofstetter P. Hearing Norton Sound: mixed methods protocol of a community randomised trial to address childhood hearing loss in rural Alaska. BMJ Open 2019 Jan 22;9(1):e023081 [FREE Full text] [doi: 10.1136/bmjopen-2018-023081] [Medline: 30674486]
70. Gauthier LV, Kane C, Borstad A, Strahl N, Uswatte G, Taub E, et al. Video Game Rehabilitation for Outpatient Stroke (VIGoROUS): protocol for a multi-center comparative effectiveness trial of in-home gamified constraint-induced movement

therapy for rehabilitation of chronic upper extremity hemiparesis. BMC Neurol 2017 Jun 08;17(1):109 [FREE Full text] [doi: 10.1186/s12883-017-0888-0] [Medline: 28595611]
71. Comparing Ways to Treat Arm Weakness Due to Stroke. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2015/comparing-ways-treat-arm-weakness-due-stroke [accessed 2021-02-24]
72. Testing a Program to Help Monitor Chronic Hepatitis B among Asian-American Patients. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2014/testing-program-help-monitor-chronic-hepatitis-b-among-asian-american-patients [accessed 2020-10-30]
73. Improving Family-Centered Pediatric Trauma Care -- The Standard of Care versus the Virtual Pediatric Trauma Center. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2020/improving-family-centered-pediatric-trauma-care-standard-care-versus-virtual [accessed 2020-12-15]
74. Margolis KL, Crain AL, Bergdall AR, Beran M, Anderson JP, Solberg LI, et al. Design of a pragmatic cluster-randomized trial comparing telehealth care and best practice clinic-based care for uncontrolled high blood pressure. Contemp Clin Trials 2020 May;92:105939. [doi: 10.1016/j.cct.2020.105939] [Medline: 31981712]
75. Comparing Transitional Care for Teens and Young Adults with Sickle Cell Disease with and without Peer Mentoring Sign Up for Updates. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2017/comparing-transitional-care-teens-and-young-adults-sickle-cell-disease-and [accessed 2020-11-11]
76. Comparing Two Ways to Help African-American and Latino Patients with HIV to Receive and Retain Care. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-two-ways-help-african-american-and-latino-patients-hiv-receive-and [accessed 2020-12-15]
77. Pekmezaris R, Nouryan CN, Schwartz R, Castillo S, Makaryus AN, Ahern D, et al. A Randomized Controlled Trial Comparing Telehealth Self-Management to Standard Outpatient Management in Underserved Black and Hispanic Patients Living with Heart Failure. Telemed J E Health 2019 Oct 01;25(10):917-925 [FREE Full text] [doi: 10.1089/tmj.2018.0219] [Medline: 30418101]
78. Pekmezaris R, Schwartz RM, Taylor TN, DiMarzio P, Nouryan CN, Murray L, et al. A qualitative analysis to optimize a telemonitoring intervention for heart failure patients from disparity communities. BMC Med Inform Decis Mak 2016 Jun 24;16:75 [FREE Full text] [doi: 10.1186/s12911-016-0300-9] [Medline: 27343060]
79. Comparing Two Ways to Help Hispanic Patients Manage Type 2 Diabetes. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-two-ways-help-hispanic-patients-manage-type-2-diabetes [accessed 2020-11-11]
80. The HOMBRE Trial: Comparing Two Innovative Approaches to Reduce Chronic Disease Risk among Latino Men. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2016/hombre-trial-comparing-two-innovative-approaches-reduce-chronic-disease-risk [accessed 2020-11-11]
81. Comparing Ways to Treat Insomnia among Black Women. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2017/comparing-ways-treat-insomnia-among-black-women [accessed 2020-11-11]
82. Comparing Two Ways to Promote Healthy Weight Gain and Prevent Obesity in Early Childhood. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-two-ways-promote-healthy-weight-gain-and-prevent-obesity-early [accessed 2020-11-13]
83. Community Health Worker and Mobile Health Programs to Help Young Adults with Sickle Cell Disease Transition to Using Adult Healthcare Services -- The COMETS Study. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2017/community-health-worker-and-mobile-health-programs-help-young-adults-sickle [accessed 2020-11-11]
84. Faro JM, Orvek EA, Blok AC, Nagawa CS, McDonald AJ, Seward G, et al. Dissemination and Effectiveness of the Peer Marketing and Messaging of a Web-Assisted Tobacco Intervention: Protocol for a Hybrid Effectiveness Trial. JMIR Res Protoc 2019 Jul 23;8(7):e14814 [FREE Full text] [doi: 10.2196/14814] [Medline: 31339104]
85. Comparing Three Ways to Help Patients Manage Multiple Chronic Conditions. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2017/comparing-three-ways-help-patients-manage-multiple-chronic-conditions [accessed 2020-12-15]
86. Nelson RG, Pankratz VS, Ghahate DM, Bobelu J, Faber T, Shah VO. Home-Based Kidney Care, Patient Activation, and Risk Factors for CKD Progression in Zuni Indians. CJASN 2018 Nov 15;13(12):1801-1809. [doi: 10.2215/cjn.06910618]
87. Shah VO, Carroll C, Mals R, Ghahate D, Bobelu J, Sandy P, et al. A Home-Based Educational Intervention Improves Patient Activation Measures and Diabetes Health Indicators among Zuni Indians. PLoS One 2015;10(5):e0125820 [FREE Full text] [doi: 10.1371/journal.pone.0125820] [Medline: 25954817]
88. Providing at-Home Kidney Disease Care for Native Americans in New Mexico. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2016/providing-home-kidney-disease-care-native-americans-new-mexico [accessed 2020-12-15]
89. Chung J, Donelan K, Macklin EA, Schwartz A, Elsharkawi I, Torres A, et al. A randomized controlled trial of an online health tool about Down syndrome. Genet Med 2021 Jan;23(1):163-173. [doi: 10.1038/s41436-020-00952-7] [Medline: 32879436]

90. Comparing Three Ways to Treat Insomnia in Adults Living in Rural Areas. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2019/comparative-effectiveness-zolpidem-and-cognitive-behavioral-therapy-insomnia [accessed 2020-12-15]
91. Tebb KP, Rodriguez F, Pollack LM, Trieu SL, Hwang L, Puffer M, et al. Assessing the effectiveness of a patient-centred computer-based clinic intervention, to reduce health disparities in unintended pregnancies among Hispanic adolescents: study protocol for a cluster randomised control trial. BMJ Open 2018 Jan 10;8(1):e018201 [FREE Full text] [doi: 10.1136/bmjopen-2017-018201] [Medline: 29326184]
92. Tebb KP, Leng Trieu S, Rico R, Renteria R, Rodriguez F, Puffer M. A Mobile Health Contraception Decision Support Intervention for Latina Adolescents: Implementation Evaluation for Use in School-Based Health Centers. JMIR Mhealth Uhealth 2019 Mar 14;7(3):e11163 [FREE Full text] [doi: 10.2196/11163] [Medline: 30869649]
93. Comparing Two Depression Prevention Programs for Teens -- The Path 2 Purpose Study. Patient-Centered Outcomes Research Institute. 2020. URL: https://www.pcori.org/research-results/2018/comparing-two-depression-prevention-programs-teens-path-2-purpose-study [accessed 2020-12-15]
94. Fernandez ME, Schlechter CR, Del Fiol G, Gibson B, Kawamoto K, Siaperas T, et al. QuitSMART Utah: an implementation study protocol for a cluster-randomized, multi-level Sequential Multiple Assignment Randomized Trial to increase Reach and Impact of tobacco cessation treatment in Community Health Centers. Implement Sci 2020 Jan 30;15(1):9-13 [FREE Full text] [doi: 10.1186/s13012-020-0967-2] [Medline: 32000812]
95. Connecting Low-Income Tobacco Users with a Telephone Quitline to Help Them Quit. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/connecting-low-income-tobacco-users-telephone-quitline-help-them-quit [accessed 2021-01-07]
96. Comparing Two Ways to Treat Childhood Obesity. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2018/comparing-two-ways-treat-childhood-obesity [accessed 2020-11-13]
97. Williams K, Markwardt S, Kearney S, Karp J, Kraemer K, Park M, et al. Addressing Implementation Challenges to Digital Care Delivery for Adults With Multiple Chronic Conditions: Stakeholder Feedback in a Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 Feb 1;9(2):e23498. [doi: 10.2196/23498]
98. Improving Cultural Competence to Reduce Health Disparities for Priority Populations. Agency for Healthcare Research and Quality. 2014. URL: https://effectivehealthcare.ahrq.gov/products/cultural-competence/research-protocol [accessed 2020-12-15]
99. Pankratz VS, Choi EE, Qeadan F, Ghahate D, Bobelu J, Nelson RG, et al. Diabetes status modifies the efficacy of home-based kidney care for Zuni Indians in a randomized controlled trial. J Diabetes Complications 2021 Feb;35(2):107753. [doi: 10.1016/j.jdiacomp.2020.107753] [Medline: 33097384]
100. Golnick C, Asay E, Provost E, Van Liere D, Bosshart C, Rounds-Riley J, et al. Innovative primary care delivery in rural Alaska: a review of patient encounters seen by community health aides. Int J Circumpolar Health 2012 Jun 29;71:18543 [FREE Full text] [doi: 10.3402/ijch.v71i0.18543] [Medline: 22765934]
101. Overview of the Alaska community health aide program. Alaska Community Health Aide Program. 2005. URL: https://www.akchap.org/resources/chap_library/Referral_Physician/CHAM_CHAP_Overview.pdf [accessed 2019-09-12]
102. Kokesh J, Ferguson AS, Patricoski C. Telehealth in Alaska: delivery of health care services from a specialist's perspective. Int J Circumpolar Health 2004 Dec;63(4):387-400. [doi: 10.3402/ijch.v63i4.17756] [Medline: 15709314]
103. Kokesh J, Ferguson AS, Patricoski C. The Alaska experience using store-and-forward telemedicine for ENT care in Alaska. Otolaryngol Clin North Am 2011 Dec;44(6):1359-74, ix. [doi: 10.1016/j.otc.2011.08.010] [Medline: 22032488]
104. Suen JJ, Bhatnagar K, Emmett SD, Marrone N, Kleindienst Robler S, Swanepoel DW, et al. Hearing care across the life course provided in the community. Bull. World Health Organ 2019 Aug 20;97(10):681-690. [doi: 10.2471/blt.18.227371]
105. PCORI Researchers Are Answering Rapidly Emerging Questions about Telehealth. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/blog/pcori-researchers-are-answering-rapidly-emerging-questions-about-telehealth [accessed 2020-12-15]
106. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2019-02-26]
107. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-12-26]
108. Arora S, Peters AL, Agy C, Menchine M. A mobile health intervention for inner city patients with poorly controlled diabetes: proof-of-concept of the TExT-MED program. Diabetes Technol Ther 2012 Jun;14(6):492-496. [doi: 10.1089/dia.2011.0252] [Medline: 22524591]
109. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
110. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med 2006 Dec;23(12):1332-1338. [doi: 10.1111/j.1464-5491.2006.01989.x] [Medline: 17116184]

111. Mulvaney SA, Anders S, Smith AK, Pittel EJ, Johnson KB. A pilot test of a tailored mobile and web-based diabetes messaging system for adolescents. J Telemed Telecare 2012 Mar;18(2):115-118 [FREE Full text] [doi: 10.1258/jtt.2011.111006] [Medline: 22383802]
112. van Olmen J, Ku GM, van Pelt M, Kalobu JC, Hen H, Darras C, et al. The effectiveness of text messages support for diabetes self-management: protocol of the TEXT4DSM study in the democratic Republic of Congo, Cambodia and the Philippines. BMC Public Health 2013 May 01;13:423 [FREE Full text] [doi: 10.1186/1471-2458-13-423] [Medline: 23635331]
113. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
114. Sieverdes JC, Treiber F, Jenkins C, Hermayer K. Improving diabetes management with mobile health technology. Am J Med Sci 2013 Apr;345(4):289-295. [doi: 10.1097/MAJ.0b013e3182896cee] [Medline: 23531961]
115. Walker RJ, Gebregziabher M, Martin-Harris B, Egede LE. Relationship between social determinants of health and processes and outcomes in adults with type 2 diabetes: validation of a conceptual framework. BMC Endocr Disord 2014 Oct 09;14(1):82 [FREE Full text] [doi: 10.1186/1472-6823-14-82] [Medline: 25298071]
116. Mehrotra A, Wang B, Snyder G. Telemedicine: What Should the Post-Pandemic Regulatory and Payment Landscape Look Like? The Commonwealth Fund. 2020. URL: https://www.commonwealthfund.org/publications/issue-briefs/2020/aug/telemedicine-post-pandemic-regulation [accessed 2021-03-09]
117. The Future Of State Telehealth Policy. National Governors Association. 2020. URL: https://www.nga.org/center/publications/the-future-of-state-telehealth-policy/ [accessed 2020-11-13]
118. Leveraging telehealth to support aging Americans. America's Health Insurance Plan. 2018. URL: https://www.ahip.org/wp-content/uploads/2018/10/AHIP-CTAC_Report.pdf [accessed 2021-03-05]

Abbreviations

CHR: community health representative
CKD: chronic kidney disease
HIPAA: Health Insurance Portability and Accountability Act
LGBT: lesbian, gay, bisexual, and transgender
MODEL: Management of Diabetes In Everyday Life
PCOR: patient-centered outcomes research
PCORI: Patient-Centered Outcomes Research Institute
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TEAM UP: Treatment Efforts Addressing Child Weight Management by Unifying Patients, Parents, and Providers
VIGoROUS: Video Game Rehabilitation for Outpatient Stroke

Edited by R Kukafka, G Eysenbach; submitted 10.03.21; peer-reviewed by MA Bahrami, H Mehdizadeh; comments to author 23.04.21; revised version received 04.06.21; accepted 03.10.21; published 07.12.21.

Please cite as:
Bailey JE, Gurgol C, Pan E, Njie S, Emmett S, Gatwood J, Gauthier L, Rosas LG, Kearney SM, Robler SK, Lawrence RH, Margolis KL, Osunkwo I, Wilfley D, Shah VO
Early Patient-Centered Outcomes Research Experience With the Use of Telehealth to Address Disparities: Scoping Review
J Med Internet Res 2021;23(12):e28503
URL: https://www.jmir.org/2021/12/e28503
doi:10.2196/28503
PMID:34878986

Review

Effectiveness of Videoconference-Delivered Cognitive Behavioral Therapy for Adults With Psychiatric Disorders: Systematic and Meta-Analytic Review

Kazuki Matsumoto[1,2], PhD; Sayo Hamatani[1,3,4], PhD; Eiji Shimizu[1,5,6], MD, PhD

[1]Research Center for Child Mental Development, Chiba University, Chiba, Japan

[2]Laboratory of Neuropsychology, Institute of Liberal Arts and Science, Kanazawa University, Kanazawa, Japan

[3]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[4]Research Center for Child Mental Development, University of Fukui, Fukui, Japan

[5]Department of Cognitive Behavioral Physiology, Graduate School of Medicine, Chiba University, Chiba, Japan

[6]Cognitive Behavioral Therapy Center, Chiba University Hospital, Chiba, Japan

Corresponding Author:
Kazuki Matsumoto, PhD
Research Center for Child Mental Development
Chiba University
1-8-1, Inohana, Chuo-ku,
Chiba
Japan
Phone: 81 43 226 2975
Fax: 81 43 226 8588
Email: axpa0219@chiba-u.jp

Abstract

Background: Cognitive behavioral therapy (CBT) is the gold standard of psychotherapy for psychiatric disorders. However, the format of delivering CBT in person limits access to the intervention. The advancements in information and communication technology, especially the internet, present an opportunity for cognitive behavioral therapists to service patients or clients in remote areas through videoconferencing. Although many randomized controlled trials of videoconference-delivered cognitive behavioral therapy (VCBT) have already been conducted, the overall estimated effect size of VCBT for psychiatric disorders has not been examined by systematic reviews and meta-analyses.

Objective: This study attempts to evaluate the effectiveness of VCBT for psychiatric disorders through a systematic and meta-analytic review.

Methods: A systematic review and meta-analysis of studies in which VCBT was directly compared to control groups (such as treatment as usual, attention control, wait-list control, and other minimal supports) was carried out. To identify previous studies that meet our study objective, 2 independent reviewers undertook a systematic search through seven databases: MEDLINE (via PubMed), Web of Science, Science Direct, PsycINFO, CINAHL, LILACS, and SciELO. Other databases (ClinicalTrials.gov and Cochrane Central Resister of Controlled Trials) were also checked. All studies included in the review were assessed using the quality criteria of the Cochrane Collaboration. Statistical analysis was performed by using Cochrane Review Manager (RevMan, version 5.4.0). Standardized mean difference was used in major meta-analyses where a *P* value of .05 or less was the threshold for statistical significance. A heterogeneity test and the chi-square test were performed to assess the presence and extent of statistical heterogeneity with significance set at $P<.10$. Funnel plots were visually inspected to assess the risk of bias. Subgroup analyses were conducted for each disorder to estimate intervention effects.

Results: The systematic search resulted in 16 studies (total N=1745) that met the criteria for this study and were included in the review. There were 10 studies on depressive symptoms, 3 on chronic pain, 1 on generalized anxiety disorder, 1 on obsessive-compulsive disorder, and 1 on hypochondriasis. The quality and risk of bias was also assessed. Results showed a pooled effect size (Hedge *g*) post treatment of −0.49 (95% CI –0.68 to –0.29), indicating that VCBT is effective for clients with psychiatric disorders. Study quality did not affect outcomes.

Conclusions: While the overall results indicate the effectiveness of VCBT, there are still only a limited number of studies on specific psychiatric and somatic conditions. Therefore, more randomized controlled trials are needed to establish the effectiveness of VCBT for different disorders.

Trial Registration: International Prospective Register of Systematic Reviews (PROSPERO) CRD42021224832; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=224832

(J Med Internet Res 2021;23(12):e31293) doi:10.2196/31293

KEYWORDS

videoconference-delivered cognitive behavioral therapy; depression; anxiety; psychiatric disorders; systematic review; meta-analysis; digital health; mental health; cognitive therapy; internet-based therapy; cognition; neurodevelopment; communication technology; health technology; psychological disorders; anxiety disorder

Introduction

Background

The incidence of mental health disorders has significant socioeconomic implications for public health and human rights globally. Depression, for example, is a leading cause of disability, affecting an estimated 264 million people worldwide [1]. It has been shown that cognitive behavioral therapy (CBT) is an effective treatment for a variety of mental disorders [2,3]. Cognitive behavioral therapists analyze the effects of their patients' cognition and behavior on psychiatric symptoms [4,5], and work toward developing adaptive cognitive-behavioral techniques with their client [6,7]. Evidence suggests that CBT is effective not only for psychiatric disorders [8] but also for somatic disorders [9]. It is considered the gold standard in the treatment of mental health disorders because it is substantiated by theory and research [10]. Further, evidence suggests that CBT is superior to other modalities, such as interpersonal psychotherapy [11]. The World Health Organization (2019) has also recognized its effectiveness and stated that access to CBT is important [12]. Face-to-face therapy is the most common format for providing treatment for mental health issues. However, this can restrict access to CBT for patients living in remote areas. Considering the widespread use of the internet and telecommunications equipment [13], there is a window of opportunity to provide access to CBT to patients living in remote areas [14]. Most remote CBT is delivered with the help of websites/webpages, under the guidance of a therapist. This format is called internet-based cognitive behavioral therapy (ICBT) or simply "internet intervention." According to Olthuis et al [15], "to be considered an Internet intervention, CBT must have been delivered over the Internet through the use of web pages or e-mail, or both." Two systematic reviews including meta-analyses have suggested that in terms of effectiveness, ICBT is equivalent to face-to-face CBT [16,17].

Videoconference-Delivered CBT

Another approach to improve accessibility to CBT for individuals residing in remote areas is to utilize a videoconferencing system [18-20]. In comparison to ICBT, videoconference-delivered CBT (VCBT) has the advantage of enabling remote treatment through interactive real-time communication between the therapist and the patient, which makes it similar to face-to-face CBT. At the same time, VCBT differs from face-to-face CBT because there is a "physical separation" between the therapist and patient, which may create limitations in clinical practice. For example, when dealing with a patient with obsessive-compulsive disorder, the patient cannot directly touch the stimulus/subject provided by the therapist. The patient must work on the subject at home on their own. Time lags and poor eye contact during video calls may affect the quality of interaction between the therapist and patient, creating obstructions for cognitive reconstruction and the creation of cognitive models.

Despite these concerns, results from previous clinical trials investigating the feasibility and efficacy of VCBT were generally promising [18-21]. The results from prospective and rigorous clinical trials (RCTs) have suggested that VCBT is not inferior to face-to-face CBT for the treatment of depression and posttraumatic stress disorder [22-25]. An existing literature review summarized previous findings [26] but did not conduct a meta-analysis of the results of RCTs drawing direct comparison to controls such as conventional treatment. A network meta-analysis that examined the most effective CBT format for the treatment of acute depression also did not compare for VCBT [27]. Several systematic reviews to assess the effectiveness of videoconference-delivered psychotherapy for the treatment of depression [28] and anxiety disorders [29] suggest that VCBT is an acceptable form of remote therapy for such patients, and clinical symptoms can be expected to improve. However, these reviews did not perform a meta-analysis owing to a lack of RCTs. Therefore, the estimate effect size of VCBT could not be gauged.

Study Objective

As of December 2020, results from new RCTs to validate the effectiveness of VCBT for people with depressive symptoms [30-35], chronic pain, and hypochondriasis [36,37] have emerged. Therefore, there is a need for a systematic review and meta-analysis focusing on VCBT. The objective of this study is to examine the effectiveness of VCBT as a treatment option for psychiatric and somatic disorders in comparison to control conditions. The population of this review targeted both clinical and community samples. To increase the credibility of the results by reviewing high-evidence studies [38], this review included only RCTs. This study's protocol was registered with PROSERO (CRD42021224832) [39]. The protocol planned to include RCTs targeting children and adolescents. However, owing to a small number of RCTs with children and adolescents as participants [40,41], the selection criteria for this review focused on studies with adult participants. This review was in accordance with the

Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement [42], see the PRISMA checklist provided in Multimedia Appendix 1.

Methods

Eligibility Criteria

The eligibility criteria for the original studies to be a part of review were as follows: (1) subjects of the study were adults (age>18 years); (2) the intervention was VCBT; (3) the intervention group was compared to attention training (AT), treatment as usual (TAU), wait-list control (WLC), or other active control (AC) conditions; (4) the outcome was the effect of VCBT on the management of the symptoms of psychiatric or somatic disorders (primary outcome measures); (5) used a randomized control study design; and (6) were written in English. Exclusion criteria were studies in which the intervention was not based on cognitive-behavioral techniques, participants were children or adolescents, and the necessary data were inaccessible.

Information Sources and Search Strategy

To identify previous studies that met our study objective, systematic searches were conducted on MEDLINE (via PubMed), Web of Science, Science Direct, PsycINFO, CINAHL, LILACS, and SciELO using the following terms related to psychiatric and somatic disorders: “depression,” “panic disorder,” “social phobia,” “social anxiety disorder,” “generalized anxiety disorder,” “obsessive-compulsive disorder,” “post-traumatic stress disorder,” “specific phobia,” “hypochondriasis,” “bulimia,” “tinnitus,” “erectile dysfunction,” “chronic pain,” or “fatigue.” To determine the intervention approach, these search terms were combined with “videoconference,” “video conference,” “videoconferencing,” “tele,” “teleconference,” “tele conference,” or “teleconferencing,” and the search filter “randomized controlled trial” was used. The search did not include unpublished studies. Searches were last updated on December 25, 2020. Other databases were also checked (ClinicalTrials.gov and Cochrane Central Resister of Controlled Trials), along with the references of the previous systematic reviews [28,29]. For more information on the full search strategies, see the complete search strategy provided in Multimedia Appendix 2.

Process of Selection and Data Collection

A total of 2 reviewers (KM and SH) independently made decisions on whether they met the selection criteria in accordance with the aforementioned search strategy. If the selected studies did not match, the decision was made by a joint discussion among the research team, including a third party (ES). The selected studies were managed using EndNote.

Data Items

Participants were those who received VCBT without restrictions, including clinical samples, community members, and students. We set the clinical symptoms as outcomes before and after the intervention. For example, if satisfaction or acceptance of the intervention was the primary outcome, a secondary outcome to measure the severity of the target disorder was adopted. The intervention was conditional on the inclusion of having sessions with the therapist via a videoconferencing system and cognitive behavioral techniques such as behavioral activation, cognitive restructuring, exposure, and mindfulness, among others.

Study Risk of Bias Assessment

The first and second authors (KM and SH) read the abstracts independently. In case of any disagreement regarding the inclusion of a particular study, the article was discussed among all researchers. All studies included in the review were assessed using the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [43,44]. The included studies were rated on each of the aforementioned dimensions as low risk, some concern, or high risk.

Statistical Analysis

Effect Measures

The standardized mean difference (SMD) was used in major meta-analyses because the studies included in this review used different symptom evaluation scales for the primary outcomes. The value of SMD depends on the effect size (difference in mean) and SD of the outcome (unique variation between participants). In the event of a missing summary statistic, we contacted the authors. If there was no reply, it was excluded from this review because there were no data that can be handled.

Synthesis Methods

Statistical analysis was performed by using Cochrane Review Manager (RevMan; version 5.4.0) [43]. First, standardization was achieved by dividing the mean difference (the change from baseline to the end of the study or the value at the end of the study) by the SD of the control group in the study. Next, in a meta-analysis, the standardized mean values from individual studies were integrated to calculate the SMD. The data reflecting intention-to-treat took precedence over the per-protocol data in the meta-analysis. Intervention effects were assessed with random-effect meta-analyses, assuming variation in true effects and accounting for the hypothesized effect distribution [45,46]. A P value of ≤.05 was considered the threshold for statistical significance. According to the power calculation by Borenstein et al [45], a power of 80% to detect a small effect size required that each group has an average of 25 participants and studies be 14 or more in number (if the probability of rejecting the null hypothesis is 5%). Subgroup analyses to the estimation of intervention effects were conducted for each disorder.

The heterogeneity test and the chi-square test were performed. Significant heterogeneity of >40% suggests the presence of heterogeneity [47]. The presence and quantity of statistical heterogeneity was assessed using the I^2 statistic, with significance set at P<.10 [48].

Sensitivity analysis was performed to compare studies that were judged to have a low risk of bias and to determine the quality of the affected outcomes.

Reporting Bias

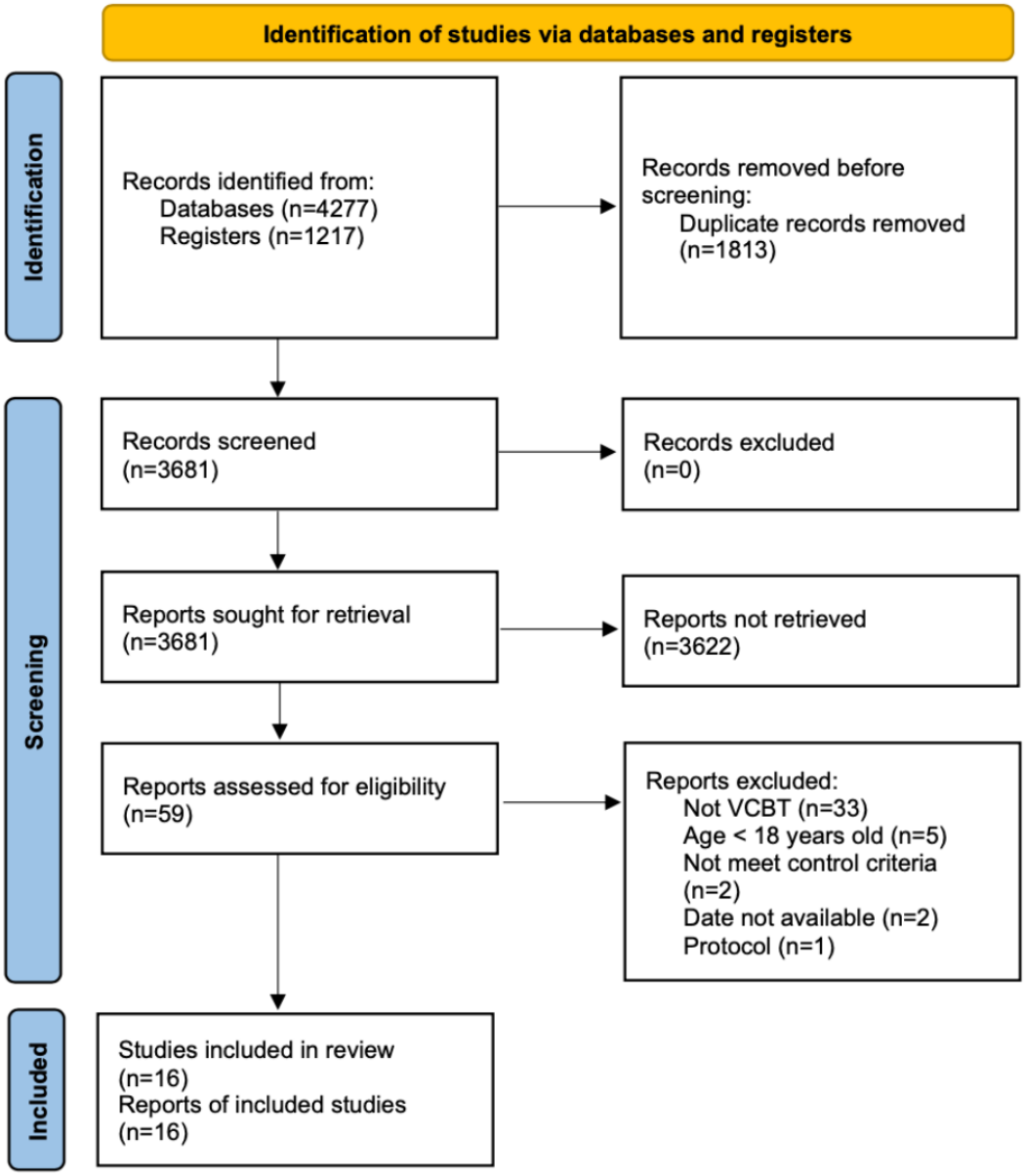
The funnel plots were visually inspected to assess the risk of bias.

Results

Studies Included in the Review

Of the 3684 screened studies, 16 (N=1745) met all selection criteria and were included in the analysis. Figure 1 shows the inclusion process based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram [49]. There was a 3-arm RCT [33]; 2 arms of those intervention groups were behavioral activation and problem-solving. In terms of participants' conditions, 10 studies targeted depressive symptoms [30,32-35,50-54], 3 focused on chronic pain [31,37,55], 1 on obsessive-compulsive disorder [21], 1 on general anxiety disorder [56], and 1 on hypochondriasis [36]. The total number of participants from whom posttreatment data were collected was 768 in VCBT and 718 in the control conditions. Studies that reported 2 RCTs were excluded because no data were available [57,58].

Figure 1. The inclusion process. VCBT: Videoconference-Delivered Cognitive Behavioral Therapy.



Characteristics of the Selected Studies

The RCTs included in this review were conducted by 13 independent research teams. Of them, 11 were performed in the United States, 2 in the United Kingdom, 2 in Canada, and 1 in Norway. The smallest study had 27 participants and the largest had 343 participants. All studies were published between 2008 and 2020. One RCT was configured with 2 VCBT intervention arms [33], the sample size of the control group set to half when the effect size was estimated in the meta-analysis. Table 1 presents the characteristics of each included study (see Multimedia Appendix 2 for details).

Table 1. Characteristics of the selected studies.

Selected studies	Diseases	Intervention		Participants, n		Outcomes	VCBT[a], mean (SD)		Control, mean (SD)		Quality	Sampling
		CBT[b] technique	Control	VCBT	Control		Pre	Post	Pre	Post		
Ahmad (2020) [30]	Depression	Mindfulness	WLC[c]	40	40	PHQ-9[d]	8.1 (6.0)	6.0 (3.9)	9.1 (6.2)	9.7 (6.9)	Low risk	Community
Alschuler (2021) [31]	Pain	CBT	TAU[e]	15	12	PCS[f]	20.1 (7.8)	15.6 (10.0)	17.4 (10.2)	17.3 (9.3)	Some concerns	Clinical
Bogosian (2015) [50]	Depression	Mindfulness	WLC	19	21	Depression HADS[g]	6.2 (3.5)	5.1 (3.2)	7.2 (3.4)	7.6 (4.0)	Low risk	Clinical
Choi (2014) [51]	Depression	PS[h]	Telephone support call	43	36	HAMD[i]	24.6 (6.6)	13.9 (7.7)	24.6 (6.6)	19.2 (7.8)	Some concerns	Community
Choi (2020a) [32]	Depression	Behavioral activation	Tele friendly visits	43	46	PHQ-9	7.2 (4.0)	5.9 (3.8)	7.7 (4.5)	8.3 (4.9)	Some concerns	Community
Choi (2020b) [33]	Depression	Behavioral activation	AC[j]	99	49	HAMD	23.2 (5.7)	14.6 (9.5)	22.9 (5.7)	18.1 (10.7)	Low risk	Community
Choi (2020b) [33]	Depression	PS	AC	98	49	HAMD	22.7 (5.7)	12.4 (10.6)	22.9 (5.7)	18.1 (10.7)	Low risk	Community
Demiris (2019) [56]	GAD	PS	AC	171	172	GAD-7[k]	6.8 (5.3)	6.2 (4.6)	7.6 (5.2)	6.6 (4.9)	High risk	Community
Elliott (2008) [53]	Depression	PS	Education only	21	14	IDD[l]	11.4 (9.4)	6.1 (6.6)	4.8 (6.1)	8.8 (13.5)	Some concerns	Community
Ferguson (2016) [54]	Depression	CBT	Supportive therapy	27	20	DASS[m] depression	6.0 (6.6)	3.7 (4.3)	12.6 (9.4)	7.3 (7.7)	Some concerns	Clinical
EI-Jawahri (2020) [52]	Depression	CBT	TAU	45	47	HADSD	4.9 (2.8)	2.9 (5.6)	3.5 (3.4)	4.4 (5.5)	Some concerns	Clinical
Fox (2020) [34]	Depression	CBSM[n]	AC	95	97	PROMIS[o] depression	49.0 (7.3)	46.6 (9.2)	48.5 (7.4)	46.6 (8.1)	Some concerns	Clinical
EI-Morr (2020) [35]	Depression	Mindfulness	WLC	80	80	PHQ-9	8.4 (5.6)	7.0 (5.0)	9.9 (6.2)	11.2 (6.7)	Low risk	Community
Morriss (2019) [36]	Hypochondriasis	CBT	TAU	78	78	SHAI[p]	24.9 (4.2)	17.7 (8.0)	25.1 (4.5)	22.6 (6.8)	Some concerns	Clinical
Somers (2018) [55]	Pain	PCST[q]	TAU	18	18	Pain severity	3.0 (2.1)	3.3 (2.4)	2.7 (1.9)	2.5 (1.9)	Some concerns	Clinical
Vogel (2014) [21]	OCD[r]	ERP[s]	WLC	10	10	Y-BOCS[t]	24.2 (4.3)	11.5 (4.8)	23.4 (2.8)	23.4 (4.8)	High risk	Clinical
Vranceanu (2019) [37]	Pain	TOR[u]	TAU	25	29	Physical function in SMFA[w]	69.8 (18.2)	20.7 (17.4)	63.2 (17.4)	48.6 (21.8)	Some concerns	Clinical

[a]VCBT: Videoconference-Delivered Cognitive Behavioral Therapy
[b]CBT: Cognitive Behavioral Therapy
[c]WLC: Wait-List Control
[d]PHQ-9: Patient Health Questionnaire, 9-item
[e]TAU: Treatment As Usual
[f]PCS: Pain Catastrophizing Scale
[g]HADS: Hospital Anxiety and Depression Scale
[h]PS: Problem Solving
[i]HAMD: Hamilton Depression Rating Scale
[j]AC: Attention Control
[k]GAD-7: Generalized Anxiety Disorder, 7-item
[l]IDD: Inventory to Diagnose Depression
[m]DASS: Depression Anxiety Stress Scales
[n]CBSM: Cognitive-Behavioral Stress Management
[o]PROMIS: Patient-Reported Outcome Measurement Information System
[p]SHAI: Short Health Anxiety Inventory
[q]PSCBT: Problem-Solving Cognitive Behavioral Therapy
[r]OCD: Obsessive-Compulsive Disorder
[s]ERP: Exposure Response Prevention
[t]Y-BOCS: Yale-Brown Obsessive-Compulsive Scale
[u]TOR: Toolkit for Optimal Recovery
[w]SMFA: Short Musculoskeletal Function Assessment Questionnaire

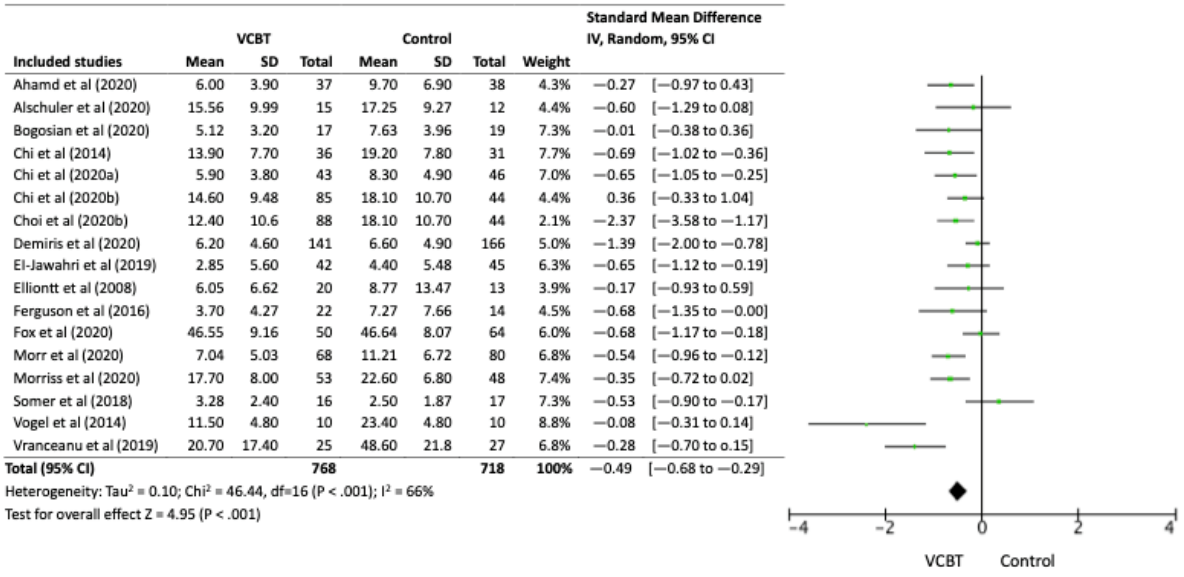
Risk of Bias in Studies

Evaluation of the quality of the studies included in this review shows that 4 RCTs (5 comparisons) are at a lower risk of bias [30,33,35,50], 10 RCTs are of some concern [31,32,34,36,37,51-55], and 2 RCTs are at a high risk of bias (see Multimedia Appendix 2 for details) [21,56].

Results of Syntheses

Figure 2, a forest plot, shows the effect size (Hedge *g*) of each study and the overall effect size integrated by the meta-analysis. An effect size estimated below 0 favors guided VCBT. In all 16 studies (17 intervention arms), the pooled between-group effect size (Hedge *g*) was −0.49 (95% CI −0.68 to −0.29, *P*<.001), showing that VCBT was significantly more effective than the control conditions.

Figure 2. Forest plot. VCBT: videoconference-delivered cognitive behavioral therapy.



In the 10 studies (11 comparing) focused on depressive symptoms, the effect size (Hedge *g*) was medium, at −0.46 (95% CI −0.60 to −0.32, *P*<.001). In 3 studies targeting chronic pain, the effect size (Hedge *g*) was −0.41 (95% CI −1.49 to 0.67, *P*=.46), showing the effectiveness of VCBT, but was not significant. In the study targeting generalized anxiety disorder, the effect size (Hedge *g*) was −0.08 (95% CI −0.31 to 0.14, *P*=.46). In the study targeting obsessive-compulsive disorder, the effect size (Hedge *g*) was −2.37 (95% CI −3.58 to −1.17, *P*<.001). In the study targeting hypochondriasis, the effect size (Hedge *g*) was −0.65 (95% CI −1.05 to −0.25, *P*=.001).

Results of the Heterogeneity Test

Tests of heterogeneity demonstrated significant differences in effects across treatments (τ^2=0.10; χ^2_{16}=46.40; I^2=66%; P<.001). The heterogeneity was largely driven by a study conducted on exposure response prevention for obsessive-compulsive disorder [21], and 2 studies that conducted unique interventions [37,55]. If those studies were excluded, the I^2 decreased from 66% to 27%: heterogeneity was not significant (τ^2=0.02; χ^2_{13}=17.87; P=.16). The pooled effect size across all studies changed marginally from Hedge g=−0.47 (95% CI −0.69 to −0.23) to Hedge g=−0.44 (95% CI −0.59 to −0.29), if those studies were excluded from the analysis. In the 10 studies (11 comparing) on depressive symptoms, tests of heterogeneity demonstrated no differences in effects across treatments (τ^2=0.01; χ^2_{10}=11.09; I^2=10%; P=.35).

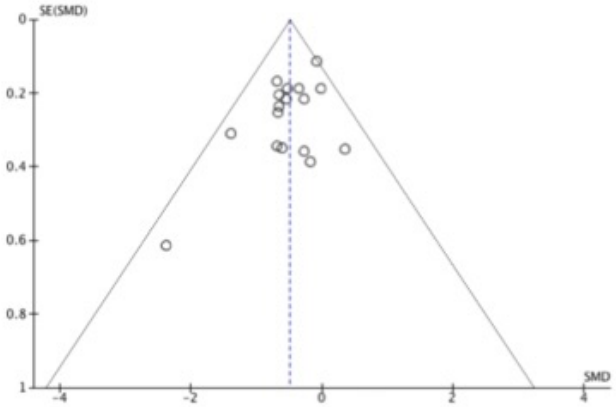
Results of Sensitivity Analysis

Subgroup analysis was conducted to verify an association between the studies' quality and intervention effects. In the 4 studies (5 comparisons) judged to have a low risk of bias, the estimated pooled effect size (Hedge g) was −0.56 (95% CI −0.74 to −0.38, P<.001; τ^2<0.001, χ^2_4=2.13, P=.71; I^2=0%) and −0.42 (95% CI −0.69 to −0.16, P=.002; τ^2=0.15, χ^2_{12}=42.01, P<.001; I^2=71%) for the other 11 studies. Thus, our results suggest that study quality did not significantly affect intervention effects.

Publication Bias

Figure 3 displays a funnel plot. Effect sizes were not evenly distributed around the averaged effect. The lower-right section of the funnel plot is devoid of studies, which suggests that there was bias in the pooled effect estimate owing to unpublished studies.

Figure 3. Funnel plot. SMD: standardized mean difference.



Discussion

Principal Findings

The objective of this systematic review was to investigate the effectiveness of VCBT as an intervention strategy as compared to control conditions such as AC, AT, TAU, and WLC using meta-analysis. Taken together, the results of this meta-analysis suggest that the pooled effect size of the primary outcomes of each disorder is medium, indicating that VCBT is especially effective for depressive symptoms. The novelty of this meta-analytic review lies in integrating the effectiveness of VCBT for psychiatric and somatic disorders, extending current knowledge into the field of remote psychotherapy [28,29,59]. However, most of the RCTs included in this review targeted depression. Therefore, further RCTs should be performed to accurately estimate the effectiveness of VCBT for generalized anxiety disorder, obsessive-compulsive disorder, chronic pain, and hypochondriasis. The quality of the study in the 16 RCTs (17 comparisons) was evaluated using RoB [43]: only 4 studies (5 comparisons) showed a low risk of bias, 10 showed some concern, and 2 showed a high risk of bias. Meta-analysis with the 4 studies (low risk of bias) revealed a moderate effect size and no heterogeneity for the effectiveness of VCBT. This result was also similar to those of a meta-analysis of 11 studies in which other judgments were made for all 16 RCTs (17 comparisons) and other bias risks. Therefore, the findings of this systematic and meta-analysis suggest evidence that VCBT is also effective for diseases for which face-to-face and guided ICBT have been demonstrated [8,16,17].

Strengths of the Review

This meta-analysis has several strengths. The results of this systematic review and meta-analysis for psychiatric disorders extend support in favor of VCBT as an effective mode of intervention. Prior systematic reviews of CBT utilizing videoconferencing systems did not include meta-analysis [28,29]. Furthermore, this study has integrated the results of RCTs that directly compared the effectiveness of VCBT for typical psychiatric disorders with minimal intervention. Our results provide useful information for clinicians and policy makers to take the practicality of VCBT into account, especially in response to the COVID-19 pandemic [60].

Limitations

This study also has some limitations. First, the analyzed studies were highly heterogeneous. Second, this study adopted a broad definition of CBT. CBT is a broad concept that includes treatment methods such as cognitive therapy, behavioral therapy, acceptance and commitment therapy, behavioral activation, problem-solving techniques, and prolonged exposure therapy. Therefore, the effect size should ideally be estimated for each therapeutic technique in the future, as was performed in a previous review of cognitive therapy for depression [61].

However, the effectiveness of each cognitive behavioral technique could not be calculated in this study, owing to the small number of pre-existing studies to analyze intervention categories and subcategories. Furthermore, owing to the inconsistent control criteria in this review, it is not possible to assess the exact effectiveness of VCBT. The gold-standard design for estimating the effect of treatment is RCT with psychological or the pill placebo group [62]. These control conditions can be standardized to control the impact of patient expectations on outcomes. Therefore, to estimate the effectiveness of VCBT accurately, it is necessary to perform a meta-analysis at the stage when RCT with psychological or pill placebo group is sufficiently accumulated. Further, this review only included studies that used adult participants owing to the low number of RCTs for children and adolescents. In future, RCTs to evaluate the effectiveness of VCBT should also be conducted on samples of children and adolescents. Since we only included articles written in English, we need to examine studies reported in more diverse languages in the future. Finally, long-term effectiveness of VCBT was not analyzed. Since this study succeeded in demonstrating the short-term effectiveness of VCBT for adult psychiatric disorders and somatic symptom disorders, the results of long-term effectiveness in the future RCTs should also be integrated.

Comparison With Prior Work

Our results, indicating that VCBT is effective for somatic disorders, extend support to those of Liu et al [9], who compared VCBT and face-to-face CBT. Additionally, the overall effect size estimated in this meta-analysis is very similar to previous results on face-to-face CBT and computerized CBT for treatment of depression and anxiety, compared to primary care TAU [61,63]. Furthermore, the effect size (Hedge g=0.46, medium effect size) of VCBT for depressive symptoms is consistent with previous results (Hedge g=0.71, large effect size) of a bias-adjusted meta-analysis [64]. Therefore, our findings demonstrate the effectiveness of VCBT as a treatment option for adults with psychiatric disorders.

This review did not include exposure techniques except for an RCT by Vogle et al [21]. Real-time interventions that utilize videoconferencing systems have the advantage of exposing the home environment of the patient [65]. At the same time, it is difficult to match up to the interpersonal experience of working with a therapist to resolve the issues, especially in the treatment for disorders that require exposure therapy such as obsessive-compulsive disorder and panic disorder. In case of obsessive-compulsive disorder, patients are afraid of things such as hospital floors, rags, and toilet paper. In case of panic disorder, the task of working with the therapist to climb stairs and exercise may be difficult to perform. Therefore, it would be premature to determine the effectiveness of VCBT for patients with anxiety disorders from the effect sizes shown in this review. The results of clinical studies that were not RCTs show that VCBT is sufficient to improve symptoms of obsessive-compulsive disorder and panic disorder [18,21,65], where exposure is an important therapeutic component [66,67]. In the future, tightly controlled RCTs are expected to be implemented, and this review should be updated with regard to the calculation of integrated estimated effect sizes for those disorders.

The number of studies included in the analysis corresponding to each condition was small, except for depressive symptoms. However, the total number of studies and participants made it possible to detect significant differences between VCBT and control conditions. However, the high degree of heterogeneity indicates the need for careful interpretation of our results. While the risk of bias detected in the quality assessment appeared to have an overall minor impact, there may be publication bias, as no negative results were reported. Namely, the funnel plot in this study suggests that there was bias in the pooled effect estimate owing to unpublished studies. Positive findings are thrice more likely to be published than negative findings [68]; hence, careful interpretation of our results is required.

Conclusions

This study attempted to provide evidence in favor of the effectiveness of VCBT as a feasible alternative approach to service patients with poor access to face-to-face CBT. VCBT, has the advantage of facilitating real-time communication between patients and therapists. This has important implications for clinicians and policy makers because it is a well-accepted approach that has demonstrated a high degree of satisfaction [19]. Although more studies are needed to draw firm conclusions, findings such as those from our meta-analysis show that VCBT is a promising treatment for future use [69].

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science KAKENHI Grant-in-aide for Scientific Research, Grant Number 18K03130 and 18K17313. We would like to thank Editage (www.editage.com) for English language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 58 KB - jmir_v23i12e31293_app1.pdf]

Multimedia Appendix 2

The complete search strategy, details of the included studies, and list of excluded studies.
[PDF File (Adobe PDF File), 1752 KB - jmir_v23i12e31293_app2.pdf]

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. Hofmann SG, Smits JAJ. Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. J Clin Psychiatry 2008 Apr;69(4):621-632 [FREE Full text] [doi: 10.4088/jcp.v69n0415] [Medline: 18363421]
3. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
4. Field T, Beeson E, Jones L. The New ABCs: A Preactitioner?s Guide to Neuroscience-Informed Cogntiive-Behavior Therapy. J Ment Health Couns 2015;37(3):A. [doi: 10.17744/1040-2861-37.3.206]
5. Hofmann SG, Asmundson GJG, Beck AT. The science of cognitive therapy. Behav Ther 2013 Jun;44(2):199-212. [doi: 10.1016/j.beth.2009.01.007] [Medline: 23611069]
6. Brewin CR. Theoretical foundations of cognitive-behavior therapy for anxiety and depression. Annu Rev Psychol 1996;47:33-57. [doi: 10.1146/annurev.psych.47.1.33] [Medline: 8624137]
7. Schacter D, Gilbert D, Wegner DM. Psychology (2nd edition). New York, NY: Worth; 2011.
8. Carpenter JK, Andrews LA, Witcraft SM, Powers MB, Smits JAJ, Hofmann SG. Cognitive behavioral therapy for anxiety and related disorders: A meta-analysis of randomized placebo-controlled trials. Depress Anxiety 2018 Jun;35(6):502-514 [FREE Full text] [doi: 10.1002/da.22728] [Medline: 29451967]
9. Liu J, Gill NS, Teodorczuk A, Li Z, Sun J. The efficacy of cognitive behavioural therapy in somatoform disorders and medically unexplained physical symptoms: A meta-analysis of randomized controlled trials. J Affect Disord 2019 Feb 15;245:98-112. [doi: 10.1016/j.jad.2018.10.114] [Medline: 30368076]
10. David D, Cristea I, Hofmann SG. Why Cognitive Behavioral Therapy Is the Current Gold Standard of Psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
11. Whiston A, Bockting CLH, Semkovska M. Towards personalising treatment: a systematic review and meta-analysis of face-to-face efficacy moderators of cognitive-behavioral therapy and interpersonal psychotherapy for major depressive disorder. Psychol Med 2019 Dec;49(16):2657-2668. [doi: 10.1017/S0033291719002812] [Medline: 31615583]
12. Mental disorders. World Health Organization. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2021-03-13]
13. Johnson J. Worldwide digital population as of October 2020. Statista, Internet, Demographics & Use 2021. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/z [accessed 2021-03-13]
14. Watts SE, Andrews G. Internet access is NOT restricted globally to high income countries: so why are evidenced based prevention and treatment programs for mental disorders so rare? Asian J Psychiatr 2014 Aug;10:71-74. [doi: 10.1016/j.ajp.2014.06.007] [Medline: 25042956]
15. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565 [FREE Full text] [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
16. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
17. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
18. Hamatani S, Numata N, Matsumoto K, Sutoh C, Ibuki H, Oshiro K, et al. Internet-Based Cognitive Behavioral Therapy via Videoconference for Patients With Bulimia Nervosa and Binge-Eating Disorder: Pilot Prospective Single-Arm Feasibility Trial. JMIR Form Res 2019 Oct 23;3(4):e15738 [FREE Full text] [doi: 10.2196/15738] [Medline: 31647472]
19. Matsumoto K, Sutoh C, Asano K, Seki Y, Urao Y, Yokoo M, et al. Internet-Based Cognitive Behavioral Therapy With Real-Time Therapist Support via Videoconference for Patients With Obsessive-Compulsive Disorder, Panic Disorder, and Social Anxiety Disorder: Pilot Single-Arm Trial. J Med Internet Res 2018 Dec 17;20(12):e12091 [FREE Full text] [doi: 10.2196/12091] [Medline: 30559094]
20. Stubbings DR, Rees CS, Roberts LD, Kane RT. Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: randomized controlled trial. J Med Internet Res 2013 Nov 19;15(11):e258 [FREE Full text] [doi: 10.2196/jmir.2564] [Medline: 24252663]

21. Vogel PA, Solem S, Hagen K, Moen EM, Launes G, Håland Å, et al. A pilot randomized controlled trial of videoconference-assisted treatment for obsessive-compulsive disorder. Behav Res Ther 2014 Dec;63:162-168. [doi: 10.1016/j.brat.2014.10.007] [Medline: 25461792]
22. Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada MA, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. Depress Anxiety 2016 May;33(5):415-423. [doi: 10.1002/da.22476] [Medline: 26864655]
23. Acierno R, Knapp R, Tuerk P, Gilmore AK, Lejuez C, Ruggiero K, et al. A non-inferiority trial of Prolonged Exposure for posttraumatic stress disorder: In person versus home-based telehealth. Behav Res Ther 2017 Feb;89:57-65 [FREE Full text] [doi: 10.1016/j.brat.2016.11.009] [Medline: 27894058]
24. Egede LE, Acierno R, Knapp RG, Lejuez C, Hernandez-Tejada M, Payne EH, et al. Psychotherapy for depression in older veterans via telemedicine: a randomised, open-label, non-inferiority trial. Lancet Psychiatry 2015 Aug;2(8):693-701. [doi: 10.1016/S2215-0366(15)00122-4] [Medline: 26249300]
25. Morland LA, Mackintosh M, Rosen CS, Willis E, Resick P, Chard K, et al. Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: randomized noninferiority trial. Depress Anxiety 2015 Nov;32(11):811-820. [doi: 10.1002/da.22397] [Medline: 26243685]
26. Thomas N, McDonald C, de Boer K, Brand RM, Nedeljkovic M, Seabrook L. Review of the current empirical literature on using videoconferencing to deliver individual psychotherapies to adults with mental health problems. Psychol Psychother 2021 Sep;94(3):854-883 [FREE Full text] [doi: 10.1111/papt.12332] [Medline: 33620133]
27. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and Acceptability of Cognitive Behavior Therapy Delivery Formats in Adults With Depression: A Network Meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
28. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing Psychotherapy and Depression: A Systematic Review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
29. Rees CS, Maclaine E. A Systematic Review of Videoconference‐Delivered Psychological Treatment for Anxiety Disorders. Australian Psychologist 2020 Nov 12;50(4):259-264. [doi: 10.1111/ap.12122]
30. Ahmad F, El Morr C, Ritvo P, Othman N, Moineddin R, Team MVC. An Eight-Week, Web-Based Mindfulness Virtual Community Intervention for Students' Mental Health: Randomized Controlled Trial. JMIR Ment Health 2020 Feb 18;7(2):e15520 [FREE Full text] [doi: 10.2196/15520] [Medline: 32074061]
31. Alschuler KN, Altman JK, Ehde DM. Feasibility and acceptability of a single-session, videoconference-delivered group intervention for pain in multiple sclerosis. Rehabil Psychol 2021 Feb;66(1):22-30. [doi: 10.1037/rep0000360] [Medline: 32790454]
32. Choi NG, Pepin R, Marti CN, Stevens CJ, Bruce ML. Improving Social Connectedness for Homebound Older Adults: Randomized Controlled Trial of Tele-Delivered Behavioral Activation Versus Tele-Delivered Friendly Visits. Am J Geriatr Psychiatry 2020 Jul;28(7):698-708. [doi: 10.1016/j.jagp.2020.02.008] [Medline: 32238297]
33. Choi NG, Marti CN, Wilson NL, Chen GJ, Sirrianni L, Hegel MT, et al. Effect of Telehealth Treatment by Lay Counselors vs by Clinicians on Depressive Symptoms Among Older Adults Who Are Homebound: A Randomized Clinical Trial. JAMA Netw Open 2020 Aug 03;3(8):e2015648 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.15648] [Medline: 32865577]
34. Fox RS, Moreno PI, Yanez B, Estabrook R, Thomas J, Bouchard LC, et al. Integrating PROMIS® computerized adaptive tests into a web-based intervention for prostate cancer. Health Psychol 2019 May;38(5):403-409 [FREE Full text] [doi: 10.1037/hea0000672] [Medline: 31045423]
35. El Morr C, Ritvo P, Ahmad F, Moineddin R, Team MVC. Correction: Effectiveness of an 8-Week Web-Based Mindfulness Virtual Community Intervention for University Students on Symptoms of Stress, Anxiety, and Depression: Randomized Controlled Trial. JMIR Ment Health 2020 Jul 17;7(7):e18595 [FREE Full text] [doi: 10.2196/18595] [Medline: 32554380]
36. Morriss R, Patel S, Malins S, Guo B, Higton F, James M, et al. Clinical and economic outcomes of remotely delivered cognitive behaviour therapy versus treatment as usual for repeat unscheduled care users with severe health anxiety: a multicentre randomised controlled trial. BMC Med 2019 Jan 23;17(1):16 [FREE Full text] [doi: 10.1186/s12916-019-1253-5] [Medline: 30670044]
37. Vranceanu A, Jacobs C, Lin A, Greenberg J, Funes CJ, Harris MB, et al. Results of a feasibility randomized controlled trial (RCT) of the Toolkit for Optimal Recovery (TOR): a live video program to prevent chronic pain in at-risk adults with orthopedic injuries. Pilot Feasibility Stud 2019;5:30 [FREE Full text] [doi: 10.1186/s40814-019-0416-7] [Medline: 30820341]
38. Greenhalgh T. How to Read a Paper: The Basics of Evidence-Based Medicine. London: BMJ Books; 2014. [doi: 10.1136/bmj.315.7112.891]
39. Matsumoto K, Hamatani S, Shimizu E. Effectiveness of remote cognitive behavioral therapy via videoconference for psychiatric and somatic disorders: a systematic review and meta-analysis. PROSPERO 2021 CRD42021224832. 2021. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021224832 [accessed 2021-06-16]
40. Wade SL, Carey J, Wolfe CR. An online family intervention to reduce parental distress following pediatric brain injury. J Consult Clin Psychol 2006 Jun;74(3):445-454. [doi: 10.1037/0022-006X.74.3.445] [Medline: 16822102]

41. Wade SL, Walz NC, Carey J, McMullen KM, Cass J, Mark E, et al. A randomized trial of teen online problem solving: efficacy in improving caregiver outcomes after brain injury. Health Psychol 2012 Nov;31(6):767-776. [doi: 10.1037/a0028440] [Medline: 22746261]
42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
43. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. London: The Cochran Collaboration; 2011.
44. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
45. Bogosian A, Chadwick P, Windgassen S, Norton S, McCrone P, Mosweu I, et al. Distress improves after mindfulness training for progressive MS: A pilot randomised trial. Mult Scler 2015 Aug 28;21(9):1184-1194 [FREE Full text] [doi: 10.1177/1352458515576261] [Medline: 25767124]
46. Crowther M, Lim W, Crowther MA. Systematic review and meta-analysis methodology. Blood 2010 Oct 28;116(17):3140-3146 [FREE Full text] [doi: 10.1182/blood-2010-05-280883] [Medline: 20656933]
47. 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]
48. 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]
49. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
50. Bogosian A, Chadwick P, Windgassen S, Norton S, McCrone P, Mosweu I, et al. Distress improves after mindfulness training for progressive MS: A pilot randomised trial. Mult Scler 2015 Aug;21(9):1184-1194. [doi: 10.1177/1352458515576261] [Medline: 25767124]
51. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
52. El-Jawahri A, Jacobs JM, Nelson AM, Traeger L, Greer JA, Nicholson S, et al. Multimodal psychosocial intervention for family caregivers of patients undergoing hematopoietic stem cell transplantation: A randomized clinical trial. Cancer 2020 Apr 15;126(8):1758-1765 [FREE Full text] [doi: 10.1002/cncr.32680] [Medline: 31899552]
53. Elliott TR, Brossart D, Berry JW, Fine P. Problem-solving training via videoconferencing for family caregivers of persons with spinal cord injuries: a randomized controlled trial. Behav Res Ther 2008 Nov;46(11):1220-1229. [doi: 10.1016/j.brat.2008.08.004] [Medline: 18828993]
54. Ferguson RJ, Sigmon ST, Pritchard AJ, LaBrie SL, Goetze RE, Fink CM, et al. A randomized trial of videoconference-delivered cognitive behavioral therapy for survivors of breast cancer with self-reported cognitive dysfunction. Cancer 2016 Jun 01;122(11):1782-1791 [FREE Full text] [doi: 10.1002/cncr.29891] [Medline: 27135464]
55. Somers TJ, Kelleher SA, Dorfman CS, Shelby RA, Fisher HM, Rowe Nichols K, et al. An mHealth Pain Coping Skills Training Intervention for Hematopoietic Stem Cell Transplantation Patients: Development and Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Mar 19;6(3):e66-e30 [FREE Full text] [doi: 10.2196/mhealth.8565] [Medline: 29555620]
56. Demiris G, Oliver DP, Washington K, Pike K. A Problem-Solving Intervention for Hospice Family Caregivers: A Randomized Clinical Trial. J Am Geriatr Soc 2019 Jul 19;67(7):1345-1352 [FREE Full text] [doi: 10.1111/jgs.15894] [Medline: 30946495]
57. Scogin F, Lichstein K, DiNapoli EA, Woosley J, Thomas SJ, LaRocca MA, et al. Effects of Integrated Telehealth-Delivered Cognitive-Behavioral Therapy for Depression and Insomnia in Rural Older Adults. J Psychother Integr 2018 Sep;28(3):292-309 [FREE Full text] [doi: 10.1037/int0000121] [Medline: 30930607]
58. Milbury K, Li Y, Durrani S, Liao Z, Tsao AS, Carmack C, et al. A Mindfulness-Based Intervention as a Supportive Care Strategy for Patients with Metastatic Non-Small Cell Lung Cancer and Their Spouses: Results of a Three-Arm Pilot Randomized Controlled Trial. Oncologist 2020 Nov;25(11):e1794-e1802 [FREE Full text] [doi: 10.1634/theoncologist.2020-0125] [Medline: 32621630]
59. Probst T, Haid B, Schimböck W, Reisinger A, Gasser M, Eichberger-Heckmann H, et al. Therapeutic interventions in in-person and remote psychotherapy: Survey with psychotherapists and patients experiencing in-person and remote psychotherapy during COVID-19. Clin Psychol Psychother 2021 Jul;28(4):988-1000 [FREE Full text] [doi: 10.1002/cpp.2553] [Medline: 33448499]
60. World Health Organization. WHO Directory-Generial's opening remarks at the media briefing on COVID-19 – 11 March 2020. World Health Organization.: World Health Organization; 2020 Mar 11. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-11-29]
61. Twomey C, O'Reilly G, Byrne M. Effectiveness of cognitive behavioural therapy for anxiety and depression in primary care: a meta-analysis. Fam Pract 2015 Mar;32(1):3-15 [FREE Full text] [doi: 10.1093/fampra/cmu060] [Medline: 25248976]

62. Lavori P. Placebo control groups in randomized treatment trials: a statistician's perspective. Biological Psychiatry 2000 Apr;47(8):717-723. [doi: 10.1016/s0006-3223(00)00838-6] [Medline: 25248976]
63. Santoft F, Axelsson E, Öst LG, Hedman-Lagerlöf M, Fust J, Hedman-Lagerlöf E. Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. Psychol Med 2019 Jun 15;49(8):1266-1274. [doi: 10.1017/S0033291718004208] [Medline: 30688184]
64. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
65. Kishimoto T, Tarumi S, Mimura M. Exposure response prevention utilizing telemedicine-case series of three cases. Japanese journal of clinical psychiatry 2016;45(12):1603-1609 [FREE Full text]
66. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: A systematic review and meta-analysis of randomised controlled trials. Compr Psychiatry 2021 Apr;106(12):152223-151609 [FREE Full text] [doi: 10.1016/j.comppsych.2021.152223] [Medline: 33618297]
67. Pompoli A, Furukawa TA, Efthimiou O, Imai H, Tajika A, Salanti G. Dismantling cognitive-behaviour therapy for panic disorder: a systematic review and component network meta-analysis. Psychol Med 2018 Sep;48(12):1945-1953 [FREE Full text] [doi: 10.1017/S0033291717003919] [Medline: 29368665]
68. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997 Sep 13;315(7109):640-645 [FREE Full text] [doi: 10.1136/bmj.315.7109.640] [Medline: 9310565]
69. Zargaran D, Walsh C, Koumpa FS, Ashraf MA, White AJ, Patel N, et al. Comment on "Internet-Based Cognitive Behavioral Therapy With Real-Time Therapist Support via Videoconference for Patients With Obsessive-Compulsive Disorder, Panic Disorder, and Social Anxiety Disorder: Pilot Single-Arm Trial". J Med Internet Res 2020 Aug 12;22(8):e13234-e13645 [FREE Full text] [doi: 10.2196/13234] [Medline: 32784172]

Abbreviations

AC: active control
AT: attention training
CBT: cognitive behavioral therapy
ICBT: internet-based cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
TAU: treatment as usual
VCBT: videoconference-delivered cognitive behavioral therapy
WLC: wait-list control

Edited by R Kukafka; submitted 22.06.21; peer-reviewed by K Matthias; comments to author 21.07.21; revised version received 14.09.21; accepted 29.10.21; published 13.12.21.

Please cite as:
Matsumoto K, Hamatani S, Shimizu E
Effectiveness of Videoconference-Delivered Cognitive Behavioral Therapy for Adults With Psychiatric Disorders: Systematic and Meta-Analytic Review
J Med Internet Res 2021;23(12):e31293
URL: https://www.jmir.org/2021/12/e31293
doi:10.2196/31293
PMID:34898445

Review

CE Accreditation and Barriers to CE Marking of Pediatric Drug Calculators for Mobile Devices: Scoping Review and Qualitative Analysis

Charlotte Koldeweij[1,2], MSc, MD; Jonathan Clarke[3], MA, BCHIR, MPH, DPhil; Joppe Nijman[4], MD, PhD; Calandra Feather[1], MRES; Saskia N de Wildt[2,5], MD, PhD, Prof Dr; Nicholas Appelbaum[1], MD, PhD

[1]Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[2]Department of Pharmacology and Toxicology, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[3]Centre for Mathematics of Precision Healthcare, Imperial College London, London, United Kingdom

[4]Department of Pediatric Intensive Care, University Medical Centre Utrecht, Utrecht, Netherlands

[5]Intensive Care and Department of Intensive Care, Erasmus MC Sophia Children's Hospital, Rotterdam, Netherlands

Corresponding Author:
Charlotte Koldeweij, MSc, MD
Department of Pharmacology and Toxicology
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein Zuid 21
Nijmegen, 6525 EZ
Netherlands
Phone: 31 243616192
Email: charlotte.koldeweij@radboudumc.nl

Abstract

Background: Pediatric drug calculators (PDCs) intended for clinical use qualify as medical devices under the Medical Device Directive and the Medical Device Regulation. The extent to which they comply with European standards on quality and safety is unknown.

Objective: This study determines the number of PDCs available as mobile apps for use in the Netherlands that bear a CE mark, and explore the factors influencing the CE marking of such devices among app developers.

Methods: A scoping review of Google Play Store and Apple App Store was conducted to identify PDCs available for download in the Netherlands. CE accreditation of the sampled apps was determined by consulting the app landing pages on app stores, by screening the United Kingdom Medicines and Healthcare products Regulatory Agency's online registry of medical devices, and by surveying app developers. The barriers to CE accreditation were also explored through a survey of app developers.

Results: Of 632 screened apps, 74 were eligible, including 60 pediatric drug dosage calculators and 14 infusion rate calculators. One app was CE marked. Of the 20 (34%) respondents to the survey, 8 considered their apps not to be medical devices based on their intent of use or functionality. Three developers had not aimed to make their app available for use in Europe. Other barriers that may explain the limited CE accreditation of sampled PDC apps included poor awareness of European regulations among developers and a lack of restrictions when placing PDCs in app stores.

Conclusions: The compliance of PDCs with European standards on medical devices is poor. This puts clinicians and their patients at risk of medical errors resulting from the largely unrestricted use of these apps.

(J Med Internet Res 2021;23(12):e31333) doi:10.2196/31333

KEYWORDS

pediatric; drug dosage calculator; European regulations; safety; medical devices; medical errors; app; application; mobile health; pharmacy

Introduction

The use of mobile health (mHealth) apps among clinicians is growing [1,2]. In 2015, 60% of medical doctors in the Netherlands used at least 1 mHealth app [3]. The widespread use of mHealth apps creates new risks for patient safety [4,5]. These risks include both technical malfunctions and misuse, either of which may lead to life-threatening medical errors [4].

To mitigate these risks, the European Union (EU)'s 2007 Medical Device Directive (93/42/ECC) (MDD) qualifies "any…software…intended by the manufacturer to be used…for the purpose of diagnosis…or treatment…of disease" as a medical device [6]. The MDD categorizes medical devices into 4 classes of risk (Classes I, IIa, IIb, and III) based on their technical characteristics, invasiveness, and potential for harm. Each class of risk determines a specific conformity assessment procedure for legally entering the European market. The higher the class of risk, the more stringent the conformity assessment procedure, with the overall objective being to provide adequate safeguards for users to be able to safely use the device. For example, for a Class II medical device, conformity assessment entails an evaluation of the device's technical documentation as well as its quality management system [6]. Depending on the device classification, conformity assessment is performed by either the manufacturer (Class I) or a European Notified Body (Class IIa and above). Once the conformity assessment is complete, medical devices obtain a CE mark, indicating their conformity with European health and safety standards, allowing them to be made available to the public within the extended single market of the European Economic Area (EEA) [7].

In May 2017, the MDD was replaced by the Medical Device Regulation (2017/745) (MDR) [8]. Among other changes, the MDR addresses software as a distinct item and establishes more stringent classification rules for software apps under Rule 11 [8,9]. By May 2021, all new devices placed on the European market were required to comply with the MDR. Devices already certified under the MDD may continue to be placed on the European market until May 2024, with the exception of Class I devices receiving a higher class under the MDR [10,11].

Despite increasingly binding European regulations, poor compliance of mHealth apps with EU certification requirements has been found. An examination of a sample of health apps freely available on several app stores by the Dutch Royal Institute for Public Health and the Environment reported that less than half are CE marked, as appropriate [12].

Pediatric drug calculators (PDCs) are tools designed to help clinicians overcome the complexities of dosing calculations in pediatrics and are increasingly used in clinical care [13]. By allowing clinicians to calculate drug doses to be administered to children based on patient characteristics, most often their weight, PDCs constitute 1 example of medical apps potentially associated with new risks for patients [14,15]. PDCs have received little scrutiny with regard to their conformity to European standards [12,16]. In this study, we perform a scoping review of Google Play Store and Apple App Store to identify PDCs available for download in the Netherlands and determine their CE accreditation status. Barriers to CE accreditation are explored through developer surveys and interviews.

Methods

Definitions

In this study, a PDC was defined as a mobile app that allows clinicians to enter information about an individual child's weight or age in order to calculate a recommended drug dosage for that child. Programs designed to determine an infusion rate or dilution volume for a given drug dosage were also defined as PDCs.

Because PDCs perform transformation of data intended to inform treatment decisions for individual patients, they would qualify as medical devices under the MDD and the MDR [6,8,17,18]. According to the MDD, PDCs would be classified as Class I medical devices [6]. In line with Rule 11 of the MDR, any software "intended to provide information…used to take decisions with…therapeutic purposes" falls under Class IIa. When such decisions can cause "a serious deterioration of a person's state of health…," the software falls under Class IIb [8]. If the decision has the potential to "cause death or an irreversible deterioration of a person's state of health," the software receives a class III classification [8]. PDCs intended for clinical use would therefore be classified as Class IIa or above under the MDR.

The terms "application provider," "manufacturer," and "developer" have been used interchangeably in this study.

Search Strategy and Screening

A scoping review of PDCs available on app stores was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [19]. Apps were searched for on Google Play Store (Android system, desktop version) and Apple App Store (iOS system, mobile version) between April 8 and April 19, 2020. Separate searches were performed in both app stores using the following search terms: "pediatric drug," "pediatric drug calculator," "neonatal drug," and "neonatal drug calculator." Sample searches for the terms "pediatric drug" and "pediatric drug calculator" were also conducted in both app stores. They produced identical results to the ones obtained for the previous search terms and were hence not completed. All sampled apps were deduplicated and screened by an individual reviewer (author CK). The availability of each app on Google Play Store and Apple App Store was verified independently of the initial search results.

Eligibility criteria were defined a priori. Apps were required to appear to be designed for health care professionals, including medical students, doctors, nurses, and paramedics. A PDC was required to be the main functionality or 1 of several functionalities of each app. The drug dosage calculator should have been developed for a pediatric population, with users able to calculate a drug dose for a specific weight, age, or body surface area. PDCs for oral or intravenous drugs were eligible if they covered more than 1 drug. Infusion dilution and infusion rate calculators were also included. Apps solely performing calculations for parenteral nutrition, maintenance fluids, electrolytes, or chemotherapy were excluded.

PDCs were screened based on their name, description, and screenshots available in each app store. Only apps that were freely available were downloaded.

Data Extraction and Qualitative Analysis

The name, manufacturer, and country of manufacture of each PDC were collected. Information about the type of calculations performed (drug dosage or infusion dilution or rate), the intended location of use (within or outside the EEA), and the number of downloads on Google Play Store were captured. To determine the CE marking status, we searched the PDC description and screenshots in app stores, any documentation provided on the app website, and relevant pages of the downloaded app (License, Disclaimer, About, or Terms and Conditions).

All PDC manufacturers with identifiable contact information were contacted through email (see Multimedia Appendix 1). Developers were invited to provide information about the type of calculations performed and the intended location of use of their PDC. They were asked whether their apps were CE marked and were invited to describe their considerations in choosing whether to pursue CE marking. They were additionally asked to report any barriers encountered in the CE accreditation process. When their responses called for clarification, they were recontacted. App providers were interviewed through video calls whenever they accepted to do so.

Data obtained from PDC manufacturers were anonymized through the attribution of a numeric code and access restricted to the first author. Responses from developers were manually analyzed. Separate considerations and barriers to CE accreditation were identified from their responses and categorized through thematic inductive analysis by 1 reviewer (CK). Developers' responses were coded against the identified themes. The coded list of barriers and considerations was discussed with 2 additional authors (JC and NA), and discrepancies were resolved through consensus.

When relevant information about CE accreditation could not be obtained from the aforementioned sources, registration of the app on the United Kingdom Medicines and Healthcare products Regulatory Agency (MHRA) was searched using the MHRA website's search function with the app manufacturer name [18]. At the time of data collection, the MHRA website constituted the only available repository of information related to the CE accreditation of medical devices in the EU.

Both app stores were contacted through their online contact pages to inquire about their review process for medical apps and the extent of their collaboration with European regulatory authorities.

Ethics

According to the Dutch Medical Research Involving Human Subjects Act, formal ethical review was not needed. Interviewees provided informed consent through email to collect and store their anonymized responses and for these to be published. Patient consent was not applicable.

Data Sharing

All data that informed this study are contained within the article and its supplementary files.

Public and Patient Involvement

Patients and the public were not involved in the design, conduct, reporting, dissemination plans of this research.

Transparency

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study, as planned, have been explained.

Results

Inclusion and Classification of Apps

A total of 632 PDCs were included for screening after deduplication (see Figure 1 and Multimedia Appendix 2). Of these, 74 (11.7%) PDCs met the inclusion criteria: 66 of the 74 (89.2%) apps were available on Google Play Store and 8 (10.8%) on Apple App Store (see Table 1); 20 (27%) apps were available on both stores. In addition, 18 of 74 (24.3%) apps were developed in EEA countries, 60 (81.1%) included a drug dosage calculator, and 14 (18.9%) incorporated an infusion rate or infusion dilution calculator without a drug dosage calculator. The number of installations per app on Google Play Store varied from 10-100 to over 100,000; 13 of 74 (17.6%) apps had been installed over 100,000 times. Of the 74 screened PDCs, only 1 (1.4%) app was CE marked.

Figure 1. Flowchart. EEA: European Economic Area.

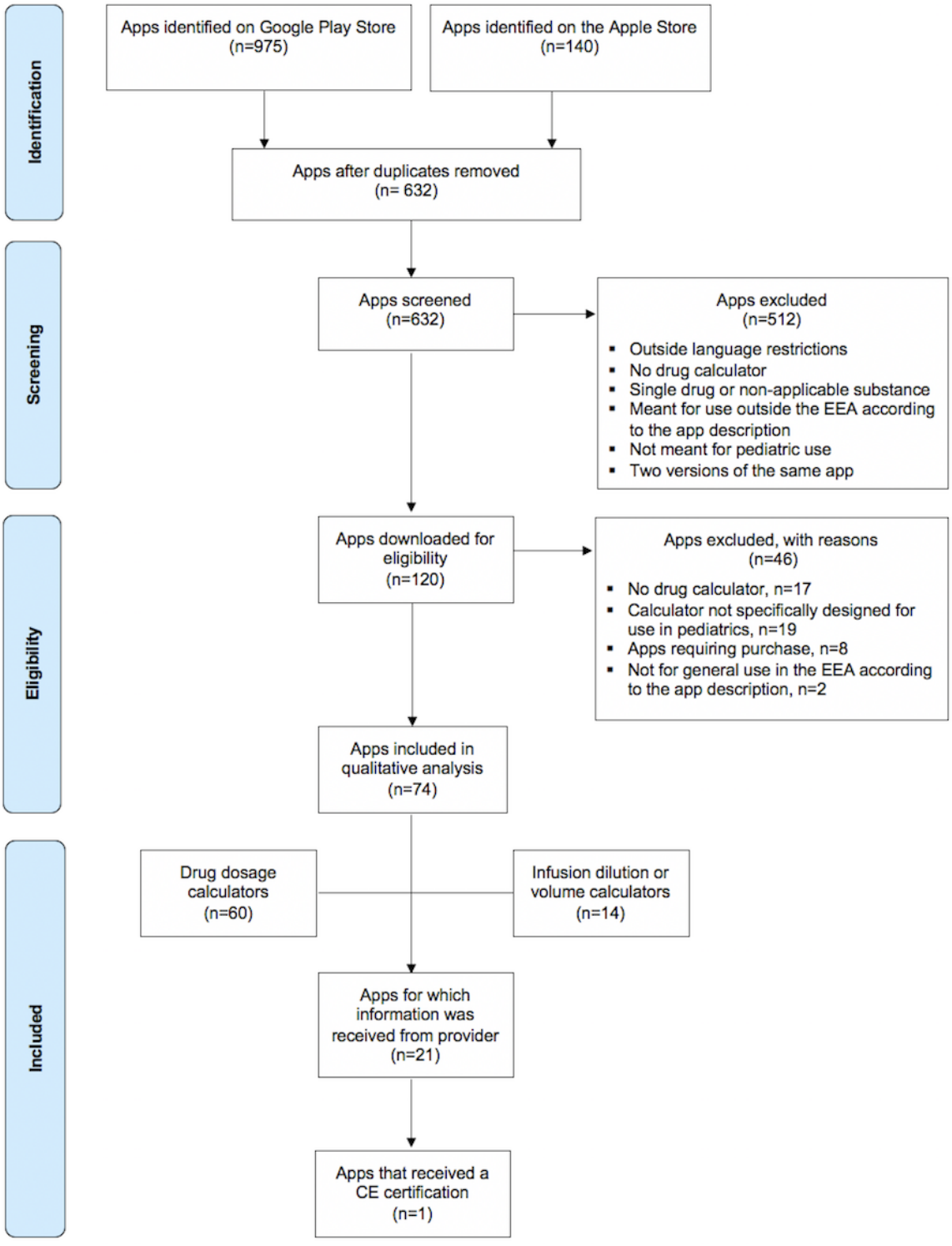


Table 1. Sampled app characteristics, including CE accreditation.

Number	App name	Country	App store	Last update[a]	Installations (n)[b]	Purchasing fee	CE marking
1	AnestCRITIC Crisis y Anestesia	Spain	Google Play	2019	5000-10,000	No	No
2	Anesthesia Assist	Portugal	Google Play	2019	50,000-100,000	No	No
3	Anesthesia Drugs Fast	United States	Google Play	2018	100-1000	Yes	No
4	Anesthesia ICC infusion calculator	Spain	Google Play	2019	5000-10,000	No	No[c]
5	Anesthesiologist	United States	Google Play	2016	Over 100,000	No	No[c]
6	Anesthetic drugs	India	Google Play	2019	5000-10,000	No	No
7	Clinical Calculator PLUS	United States	Google Play	2020	1000-5000	Yes	No
8	CoPE Paediatric Emergency	Denmark	Google Play	2019	1000-5000	Yes	No
9	Dosage Calculator	Hong Kong	Google Play	2019	Over 100,000	Yes	No
10	Dose calculator	Egypt	Google Play	2020	Over 100,000	No	No
11	Dosefinder 1	United Kingdom	Google Play	2016	5000-10,000	No	No[c]
12	Dosis Pediatricas	__[d]	Google Play	2019	1000-5000	No	No
13	DosisPedia	Spain	Google Play, Apple App	2020	Over 100,000	No	No[c]
14	DrDrugs: Drug Guide for Physicians - 2020 Updates	United States	Google Play	2020	1000-5000	Yes	No
15	Drug dosage calculations	Saudi Arabia	Google Play	2018	50,000-100,000	No	No[c]
16	Drug Dose	Ukraine	Google Play	2016	5000-10,000	No	No
17	DrugCalc: Pediatric dosing calculator	Thailand	Google Play	2017	5000-10,000	No	No[c]
18	DrugDoses	United States	Google Play, Apple App	2019	5000-10,000	Yes	No[c]
19	Drugscape dose calculator	Jordan	Google Play	2019	10,000-50,000	No	No
20	Easy Drug Dose Calculator	Australia	Google Play	2018	Over 100,000	No	No
21	EBMcalc Pediatrics	United States	Apple App	—	—	Yes	No[c]
22	eBroselow SafeDose	United States	Google Play, Apple App	2020	Over 100,000	No	No
23	EMS Calculator	United States	Apple App	—	—	Yes	No
24	EMS Drugs Fast	United States	Apple App	—	—	Yes	No
25	EnfermerApp	Chile	Google Play	2019	10,000-50,000	No	No[c]
26	GIR Calc	United States	Apple App	—	—	No	No
27	Infinite dose: the smart dosage calculator	Egypt	Google Play	2018	10,000-50,000	No	No[c]
28	Infusions	Colombia	Google Play, Apple App	2020	Over 100,000	No	No
29	Infusions - Infusions Calculator	Egypt	Google Play	2019	10,000-50,000	No	No[c]

Number	App name	Country	App store	Last update[a]	Installations (n)[b]	Purchasing fee	CE marking
30	Inotropes Rate Calculator	Jordan	Google Play	2016	5000-10,000	No	No[c]
31	Intravenous Medications Gahart	United States	Google Play	2019	5000-10,000	Yes	No[c]
32	Kids Drug Dosage Calc - PaedRx	India	Google Play	2013	50,000-100,000	No	No
33	Lexicomp	United States	Google Play, Apple App	2020	Over 100,000	Yes	No[c]
34	Medic Dose Calculator	India	Google Play	2018	5000-10,000	No	No[c]
35	Medical Calculator	United States	Google Play	2019	500-1000	No	No[c]
36	Mediquations Medical Calculator	United States	Google Play, Apple App	2018	10,000-50,000	Yes	No
37	Millidos: Pediatric Drug Dosages	Syria	Google Play	2019	10,000-50,000	No	No
38	MKD Dosage Calc	—	Google Play	2019	100-1000	No	No
39	Neomate	United Kingdom	Google Play, Apple App	2017	50,000-100,000	No	Yes
40	NeonaCal	Ireland	Apple App	—	—	Yes	No
41	Neonatology	United Kingdom	Google Play, Apple App	2019	100-1000	Yes	No
42	NICU	United Kingdom	Google Play, Apple App	2019	100-1000	Yes	No
43	Nursing calculator	India	Google Play	2020	Over 100,000	No	No[c]
44	Paediatric Emergencies	United Kingdom	Google Play	2019	1000-5000	No	No
45	Paediatric Emergency Tools	United Kingdom	Google Play, Apple App	2019	100-1000	Yes	No
46	palmPEDi: Pediatric Tape	United States	Google Play, Apple App	2013	5000-10,000	Yes	No
47	Paramedic Meds	United States	Google Play	2019	10,000-50,000	Yes	No
48	PedAMINES	Switzerland	Google Play, Apple App	2018	10-50	Yes	No
49	Ped(z) - Pediatric Calculator	Germany	Google Play, Apple App	2017	Over 100,000	No	No[c]
50	PedCalc	Egypt	Google Play	2017	10,000-50,000	No	No[c]
51	Pedi Crisis 2.0	United States	Google Play, Apple App	2019	1000-10,000	No	No
52	Pedi Help	Switzerland	Google Play, Apple App	2017	50,000-10,000	No	No
53	Pedi Safe Medications	United States	Google Play	2016	10,000-50,000	Yes	No
54	Pedi Safe Pediatric Anesthesia	United States	Apple App	—	—	No	No
55	Pedi STAT	Canada	Google Play, Apple App	2018	Over 100,000	Yes	No
56	Pediatria calculadora dosis/kg	—	Google Play	2020	Over 100,000	Yes	No
57	Pediatric dosage calculator	Hong Kong	Google Play	2019	Over 100,000	No	No

Number	App name	Country	App store	Last update[a]	Installations (n)[b]	Purchasing fee	CE marking
58	Pediatric dose calculator	Netherlands	Google Play, Apple App	2016	50-100	Yes	No
59	Pediatric doses calculator	Egypt	Google Play	2020	10,000-50,000	No	No
60	Pediatric Gas for Anesthesia	United States	Apple App	—	—	Yes	No
61	Pediatric Guideline/Emergency/Pediatric child care	United Kingdom	Google Play	2020	Over 100,000	No	No
62	Pediatric IV calculator	Netherlands	Google Play, Apple App	2016	10-50	Yes	No
63	Pediatric IV dosage	—	Google Play	2014	50,000-100,000	No	No
64	Pediatric IV Rate	United States	Google Play	2019	50-100	No	No[c]
65	Pediatric oral dosage	—	Google Play	2015	50,000-100,000	No	No
66	Pediatric pedia	Middle East	Apple App	—	—	No	No
67	PediRef: Pocket Pediatrics	United States	Google Play	2017	10,000-50,000	No	No
68	PedsGuide	United States	Google Play, Apple App	2019	1000-5000	No	No
69	PeKemecum	Spain	Google Play	2019	50,000-100,000	No	No
70	PICU Calculator	United Kingdom	Google Play, Apple App	2019	1000-5000	No	No
71	PICUDoctor 5 - Cardiac Guide	Australia	Google Play	2015	10,000-50,000	Yes	No[c]
72	RightDose	United States	Apple App	—	—	No	No
73	SmartPedi-Pediatric Treatment & Dose Calculator	Bangladesh	Google Play	2019	5000-10,000	No	No
74	UCIN-Calc Beta	Dominican Republic	Google Play	2018	5000-10,000	No	No

[a]Last update on Google Play Store.

[b]Number of installs on Google Play Store on May 8, 2020.

[c]Information obtained from the app developer.

[d]Not available.

Qualitative Analysis

App Developers

Of 61 app developers, 59 (96.7%) for whom contact information was available were contacted through email; 1 (1.6%) developer was additionally contacted through a video call. Responses were obtained from 20 of 59 (33.9%) providers that developed 21 apps (see Table 2). Of the 20 developers, 3 (15%) were based in the EEA. None of the apps developed by the respondents were CE marked. In addition, 2 of the 20 developers (10%) indicated that they understood that their apps qualified as Class I medical devices under the MDD but were not CE marked (developers 4 and 8), while 2 (10%) had attempted to get their apps CE marked but were unsuccessful (developers 5 and 18).

The most frequent reason for not pursuing CE accreditation provided by developers was that in their view, their apps did not qualify as medical devices (8/20, 40%). Various arguments informed this assessment. Of the 20 developers, 2 (10%) referred to the intended use of their apps, stating that the apps were designed as a *reference* or an *educational* tool for clinicians as opposed to a clinical decision-making aid (developers 11 and 15). This disclaimer was also frequently provided in the end-user licenses of sampled apps. Other developers referred to their apps' functionality, describing them as *digital documents* (developer 1) or *books* (developer 14), which did not entail manipulation of data. In both cases, the functionality of the apps involved transformation of data. Arguments pertaining to functionality also examined the nature of the information being input into and delivered by a given app, and the weight of the result in determining the process of care. Developer 18 highlighted a difference between drug dosage calculators that could be seen as medical devices owing to their recommending a specific drug dose based on an individual patient's characteristics, and the infusion rate or dilution calculators performing simple conversion operations on pre-established drug prescriptions. According to developers 8 and 13, the level of transparency and complexity of the computations performed

by an app constituted key factors when determining whether it qualified as a medical device. Developer 13 suggested that if the calculations performed by an app are simple enough to be immediately replicable by users, then the app would not qualify as a medical device. Developer 8 suggested that even for more complex calculations, if the calculations are linked to user-accessible formulae and bibliographic support, the app should not be classified as a medical device.

Other reasons put forward by developers as justification for not CE-marking their apps included a lack of knowledge of European legislation on medical devices (3/20, 15%), the fact that their apps were not devised for use in EEA countries (3/20, 15%), and the fact that no certification was required for access to Google Play or Apple App Store (4/20, 20%). Several manufacturers described app stores as implicit arbiters for matters of regulatory compliance or safety ("My app was evaluated in the…store by the public user" or "It was very easy to place on the…store."). Of the 20 providers, 3 (15%) indicated that Apple App Store was more restrictive than Google Play Store when granting access for PDCs; 1 (5%) developer outside the EEA argued that his app did not require testing or accreditation according to the regulations of his country (developer 15).

Several barriers to CE marking were outlined by developers. Of the 20 manufacturers, 2 (10%) indicated that the process was too complex (developers 4 and 18), and 1 (5%) said it was too costly to take on as an individual developer or a small enterprise (developer 18). This appeared more generally relevant across the sample, with multiple developers stating that they were clinicians with programming skills who developed a PDC "as a hobby" (developer 6) or "for their own use" (developer 16). An added barrier in this view concerned the lack of institutional support received by app manufacturers seeking to obtain a CE marking that were also affiliated to a hospital or a university. After receiving confirmation from national regulatory authorities that his app qualified as a Class I medical device under the MDD, developer 4 asked the relevant national health care institution for its support in the CE accreditation process. He did not obtain this support due to the institution's concerns over the costs and associated legal liability. He shared that "developers are often left unsupported by their associated institutions...I think mostly because of a lack of experience and knowledge regarding the governance and legal implications, many institutions feel vulnerable and unwilling to engage with regulatory bodies." Overall, this "had an unfortunate regressive effect" on the dissemination of his app. Independently of CE accreditation, 5 of 20 (25%) developers had sought alternative forms of clinical validation for their apps, for example, by national experts.

Table 2. Developer responses on the barriers to CE accreditation.

Barriers to CE accreditation and other considerations outlined by developers on the CE accreditation process	Developer[a]																				Total developers (n)
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
No reason provided	✗[b]	✓[c]	✗	✗	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	5
App not meant for use in European countries	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	3
Not a medical device	✓	✗	✗	✗	✗	✗	✗	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗	✓	✗	✓	8
Unaware of the EEA[d] medical device regulations	✗	✗	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	3
Compliant with national regulations (non-EEA)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	1
App store not requiring certification	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	3
Discussed with national certification authorities	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	1
Did not receive institutional support	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	1
Process too complex	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	2
Process too costly	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	1
App undergoing another form of validation	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	5

[a]Of the 20 providers, 1 (5%) had developed 2 apps; we did not indicate which one in order to prevent its identification.

[b]✗: no.

[c]✓: developer provided this specific reason.

[d]EEA: European Economic Area.

App Stores

Neither store provided information about its review process and collaboration with European regulatory authorities. Apple App Store's guidance states that "drug dosage calculators must come from the drug manufacturer, a hospital, university, health insurance company, pharmacy or other approved entity, or receive approval by the FDA or 1 of its international

counterparts" [20]. No such clause was found in the Google Play Store guidance [21].

Discussion

Principal Findings

We systematically reviewed the CE accreditation of PDCs available on 2 mobile app stores in the Netherlands. Of 74 sampled PDCs, 1 (1.4%) had the appropriate CE marking in conformity with the MDD. At a time when European regulatory authorities are seeking to enhance their scrutiny of medical apps, for example, through the MDR, this study sheds a new light on several barriers to CE accreditation for eligible mHealth apps.

This study delivered several new insights. It revealed that almost all PDCs available for download in the Netherlands fail to comply with European regulations on medical devices. The only app that is certified under the MDD (Neomate; see Table 1) will likely require additional assessment due to the more stringent classification requirements of the MDR [3,8,22]. The status quo with regard to CE accreditation for PDCs available on app stores is concerning, especially considering the fact that these apps are widely used by clinicians and have the potential to cause harm. Of the 74 PDCs identified on the screened stores, 13 (17.6%) had been downloaded over 100,000 times. Our findings thus echoed those of earlier studies highlighting the widespread use of mHealth apps among clinicians, including pediatricians [1,3].

Multiple reasons were identified for PDC manufacturers' poor compliance with European regulations. First, PDC developers appeared to have varying levels of awareness of the existence of such regulations. For those manufacturers that knew about these regulations, European rule interpretation was ambivalent. Several developers argued that their apps do not qualify as medical devices according to the relevant European standards. This was true despite the clear statement by the MDD, the MDR, and associated European and European member state guidance that any software involving manipulation of data intended to be used for diagnostic or treatment purposes in individual patients qualify as a medical device [3,6,8,17,22]. The concept of *intent of use* seemed especially prone to a variety of interpretations by manufacturers. Many of those interviewed, as well as the end-user licenses of multiple sampled apps, indicated that their PDCs were for reference or educational purposes only. This claim, however may be in conflict with the *actual* use of such apps by their users, given their functionality. Although data on PDC usage by clinicians is scarce, anecdotal evidence suggests that the advice generated by such apps is frequently used to inform real patient care.

Reflecting on the functionality of their apps, some developers highlighted a difference between pediatric drug dosage calculators and calculators of infusion volumes or rates [23]. The difference, they contended, pertained to the type of information being input into the app and the data delivered by it, as well as the complexity and transparency of the computations it ran. Although drug dosage calculators generated medication advice based on individual patient characteristics, this was not the case for infusion rate calculators that performed conversion calculations on a pre-established drug dosage. This distinction, however, does not align with MDD guidance nor with the MDR, which take the stance that any app involving transformation of data subsequently informing the treatment of an individual patient qualifies as a medical device, irrespective of the complexity of the transformation [6,7,17].

In addition to disagreements on the substance of European law, another barrier hampering broader CE accreditation of eligible apps concerned the technical nature and potential costs associated with this process. According to the MDD and the MDR [6,8], the onus of certification falls on providers that may lack the capacity to take on the associated liability and costs [24,25]. The challenging nature of the conformity assessment process will only increase under the MDR, given the up-classification of software apps, leading to additional evaluation requirements, including the appointment of a notified body [8,9]. In this context, a general lack of institutional support for developers seeking CE accreditation for their apps may become even more discouraging.

Another factor likely to undermine the compliance of PDC manufacturers with European standards on medical devices concerns the lack of an established process for enforcing these rules at a premarket stage. As with other European legislations, the enforcement of the MDD and the MDR is incumbent upon each EU member state [26]. Although the Dutch Decision on Medical Devices states that the distribution and use of apps that fail to obtain a CE mark is forbidden [27], it does not provide any enforcement means before such apps become available on app stores. Restrictive measures are unlikely to be taken unless a medical error resulting from the use of software occurs, especially if the latter leads to litigation. In this case, the responsibility for the medical error falls on both app users and the app developer [3]. Although the MDR tightens the requirements for CE accreditation and enhances postmarket surveillance [28], it does not fundamentally change the principle by which software manufacturers are themselves responsible for initiating the CE marking process [8]. Effectively, the EU's reliance on this framework in the absence of institutional support for developers and of control mechanisms at a premarket stage may have contributed to making other actors, for example, app stores, informally responsible for restricting European market access. It also implies that clinicians (or their institutions) wishing to use a PDC should themselves assess whether an app is properly accredited despite their lack of expertise in such matters [5].

Among other measures, these findings speak to the need for making CE marking information more readily available to PDC users. This may be achieved through the planned extension of the European Database on Medical Devices (EUDAMED), scheduled to become publicly accessible in May 2022 [29,30], and through the introduction of unique device identifiers for medical devices across the EEA, expected by 2024 [31]. EEA member states may also choose to build on existing online registries of certified or evidence-based apps [16,32]. European authorities could seek to formally engage app stores as partners in the enforcement of the European MDR. At this stage, it appears that the initiative for restricting access to app stores resides with the app stores themselves, as illustrated by the

various levels of restrictions described in the guidance documents of Apple App Store and Google Play Store. The finding that more PDCs were available on Google Play Store (66/74, 89.2%) than on Apple App Store (28/74, 10.8%) may suggest that differences in the stringency of requirements contributed to developers' decisions on where to make their apps available.

Limitations

This work had several limitations. Web-based PDCs that did not have a mobile interface, for example, the Dutch Paediatric Formulary calculator, which was developed in conformity with the requirements of the MDD [33,34], were excluded. The restricted search functions of app stores limited the comprehensiveness of the search possible, for example, excluding paid-for apps. As a result, the list of PDCs included from those stores may not be exhaustive and may only apply to apps available for download in the Netherlands. Eight apps that were only available for purchase were excluded. Considering the potential differences between freely available apps and apps that were available for purchase and whose manufacturers may thus rely on additional finances to recover the costs associated with obtaining a CE marking, this could have led to selection bias. Despite the existence of MDD guidance stating that CE accreditation should be clearly displayed on app landing pages in the relevant stores [17] and our cross-referencing of multiple sources, it is possible that 1 or more CE-marked PDCs were misclassified. In the absence of a mandatory statement on CE accreditation on the app stores, PDC developers were contacted directly. Additionally, 2 of the 61 (3.3%) developers could not be contacted due to missing contact information, and only 20 of the 59 (33.9%) developers contacted provided responses. This relatively low response rate was likely to introduce response bias into the qualitative component of the study. We expect therefore that those developers who responded may represent those who wish to be accessible to those with questions about their apps, and as such their responses may not be representative of all app developers.

Conclusion

This study demonstrates that almost no PDC currently available on two app stores accessed in the Netherlands adheres to European regulations on CE marking. In addition to the limited awareness of these norms among PDC developers, this compliance gap can be related to incorrect rule interpretation by some app manufacturers, the lack of mechanisms for verifying mHealth apps' compliance with European medical device rules before market access, and the technical nature of the CE accreditation process for developers often lacking institutional support.

Although limited to a single category of apps, it is likely that these findings apply to a broader set of mobile devices being used in clinical settings. This lack of regulatory compliance puts both clinicians and patients at risk of medical errors resulting from the use of uncertified and, in some cases, potentially unsafe PDCs. This practice therefore undermines the potential impact of the MDD and the MDR, which strive to create a technologically safer European medical landscape, while supporting clinicians' trust in the devices they use.

Acknowledgments

This work was supported by the National Institute for Health Research and the Wellcome Trust (Sir Henry Wellcome Fellowship).

Authors' Contributions

CK collected and analyzed data and drafted and revised the paper. She is the guarantor. NA conceptualized the paper, together with CK. NA, JC, JN, and SW revised the paper. CK also attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

The authors had no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and had no other relationships or activities that could appear to have influenced the submitted work. NA is director of Dosium Holdings Ltd, a software company developing computerized decision support tools for medication safety. SW is director of the Dutch Pediatric Formulary and is a paid consultant for AMPharma, Khondrion. SW receives research funding from the EU (IMI JU2) and the Bill & Melinda Gates Foundation. CF is a shareholder in Dosium Holdings Ltd. The study funders were not involved in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The lead author confirms the independence of the contributing researchers from funders and that all authors had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Multimedia Appendix 1
Email to developers (email body).
[DOCX File , 14 KB - jmir_v23i12e31333_app1.docx]

Multimedia Appendix 2
List of excluded apps and reasons for exclusion.
[DOCX File , 19 KB - jmir_v23i12e31333_app2.docx]

References

1. Hadjipanayis A, Klonis A, Abdel-Mannan OA, Sutcliffe AG. Smartphone use in paediatric practice: a national survey. J Mob Technol Med 2016 Mar;5(1):3-8. [doi: 10.7309/jmtm.5.1.2]
2. 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 1;7(11):e15195-e15112. [doi: 10.2196/15195]
3. Ekker A, van Rest B. Medische apps, is certificeren nodig?. URL: https://www.nictiz.nl/wp-content/uploads/2013/05/13005-Whitepaper-medische-apps.pdf [accessed 2013-06-21]
4. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
5. Morse SS, Murugiah MK, Soh YC, Wong TW, Ming LC. Mobile health applications for pediatric care: review and comparison. Ther Innov Regul Sci 2018 May;52(3):383-391. [doi: 10.1177/2168479017725557] [Medline: 29714538]
6. European Council. Directive 93/42/EEC of 14 June 1993 concerning medical devices. Off J Eur Union 1993;OJ L 169, 12:1-43.
7. European Commission. CE Marking. URL: https://ec.europa.eu/growth/single-market/ce-marking_en [accessed 2021-06-13]
8. European Parliament, European Council. Regulation 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance). Off J Eur Union 2017 May 05;OJ L 117:1-175.
9. Martelli N, Eskenazy D, Déan C, Pineau J, Prognon P, Chatellier G, et al. New European regulation for medical devices: what is changing? Cardiovasc Intervent Radiol 2019 Sep;42(9):1272-1278. [doi: 10.1007/s00270-019-02247-0] [Medline: 31123774]
10. European Commission. Transition Timelines from the Directives to the Regulations Medical Devices and In Vitro Diagnostic Medical Devices. URL: https://ec.europa.eu/docsroom/documents/34907 [accessed 2021-06-13]
11. European Commission. Commission Postpones Application of the Medical Devices Regulation to Prioritise the Fight against Coronavirus. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_589 [accessed 2021-06-13]
12. van Drongelen A, de Bruijn A, Roszek B, Vonk R. Apps under the medical devices legislation. National Institute for Public Health and the Environment. 2019. URL: https://rivm.openrepository.com/handle/10029/622397 [accessed 2021-06-13]
13. Murray B, Streitz MJ, Hilliard M, Maddry JK. Evaluation of an electronic dosing calculator to reduce pediatric medication errors. Clin Pediatr (Phila) 2019 Jan 08;58(4):413-416. [doi: 10.1177/0009922818821871]
14. Hirata K, Kang A, Ramirez G, Kimata C, Yamamoto L. Pediatric weight errors and resultant medication dosing errors in the emergency department. Pediatr Emerg Care 2019;35:637-642. [doi: 10.1097/pec.0000000000001277]
15. Tse Y, Tuthill D. Incidence of paediatric 10-fold medication errors in Wales. Arch Dis Child 2021 Jul;106(7):656-661. [doi: 10.1136/archdischild-2020-319130] [Medline: 33109522]
16. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3(1):14-16 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
17. European Commission. Guidelines on the Qualification and Classification of Stand Alone Software Used in Healthcare within the Regulatory Framework of Medical Devices. URL: https://ec.europa.eu/health/sites/default/files/md_topics-interest/docs/md_meddev-guidance-216_en.pdf [accessed 2021-12-02]
18. Medicines and Health Products Regulatory Agency. Public Access Database for Medical Device Registration. URL: https://aic.mhra.gov.uk/era/pdr.nsf/name?openpage&start=7601&count=200 [accessed 2021-06-13]
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
20. Apple. App Store Review Guidelines Internet. URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2021-01-12]
21. Google. Play Console Help. URL: https://support.google.com/googleplay/android-developer/?hl=en#topic=7071529 [accessed 2021-01-12]
22. Medicines and Healthcare products Regulatory Agency. Guidance: Medical Device Stand-Alone Software Including Apps (Including IVDMDs) v1.05. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957090/Software_flow_chart_Ed_1-07b-UKCA__002__FINAL.pdf [accessed 2021-06-13]

23. Ferreras Vega R. El uso de aplicaciones móviles en medicina. Rev Electron AnestesiaR 2018;10:1-3 [FREE Full text]
24. Maresova P, Hajek L, Krejcar O, Storek M, Kuca K. New regulations on medical devices in Europe: are they an opportunity for growth? Adm Sci 2020;10:1-18. [doi: 10.3390/admsci10010016]
25. European Commission. Internal Market, Industry, Entrepreneurship and SMEs. URL: https://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en [accessed 2021-06-13]
26. CAMD Europe. Competent Authorities for Medical Devices. Implementation of the new EU Medical Devices Regulations. URL: https://www.camd-europe.eu/mdr-ivdr-implementation/ [accessed 2021-06-13]
27. Overheid Wettenbank. Besluit Medische Hulpmiddelen. URL: http://wetten.overheid.nl/BWBR0007307/geldigheidsdatum_23-03-2013 [accessed 2021-06-13]
28. Behan R, Watson M, Pandit A. New EU medical device regulations: impact on the MedTech sector. Med Writ 2017;26:20-24 [FREE Full text]
29. European Commission. European Database on Medical Devices (EUDAMED). URL: https://ec.europa.eu/growth/sectors/medical-devices/new-regulations/eudamed_en [accessed 2021-06-13]
30. Fraser AG, Butchart EG, Szymánski P, Caiani EG, Crosby S, Kearney P, et al. The need for transparency of clinical evidence for medical devices in Europe. Lancet 2018;392:521-530.
31. European Commission. Unique Device Identification System under the EU Medical Device Regulations 2017/745 and 2017/746. URL: https://ec.europa.eu/health/sites/default/files/md_topics-interest/docs/md_faq_udi_en.pdf [accessed 2021-06-13]
32. Arts Portaal. Het overzicht van apps voor zorgprofessionals. URL: https://www.artsportaal.nl/apps/ [accessed 2021-06-13]
33. Kinderformularium. Kinderformularium.nl. URL: https://www.kinderformularium.nl/ [accessed 2021-06-13]
34. van der Zanden TM, Goedknegt L, De Hoog M, Mooij MG, de Wildt SN, van der Sijs IH. Development and implementation of a paediatric dosing calculator integrated in the Dutch Paediatric Formulary. Drugs Ther Perpect 2020:253-262.

Abbreviations

EEA: European Economic Area
EU: European Union
EUDAMED: European Database on Medical Devices
MDD: Medical Devices Directive
MDR: Medical Device Regulation
MHRA: Medicines and Healthcare products Regulatory Agency
PDC: pediatric drug calculator
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CE: Conformité Européenne

Edited by R Kukafka; submitted 22.07.21; peer-reviewed by B Nievas Soriano, C Jacob; comments to author 09.10.21; revised version received 01.11.21; accepted 02.11.21; published 13.12.21.

Please cite as:
Koldeweij C, Clarke J, Nijman J, Feather C, de Wildt SN, Appelbaum N
CE Accreditation and Barriers to CE Marking of Pediatric Drug Calculators for Mobile Devices: Scoping Review and Qualitative Analysis
J Med Internet Res 2021;23(12):e31333
URL: https://www.jmir.org/2021/12/e31333
doi:10.2196/31333
PMID:34898456

Review

Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis

Lindsay H Dewa[1,2*], BA, MSc, PhD; Emma Lawrance[1,3*], BSc, Grad Dip (Science Communication), MSc, DPhil; Lily Roberts[1], BSc, PGCert; Ellie Brooks-Hall; Hutan Ashrafian[1], BSc, MBBS, MRCS, MBA, PhD; Gianluca Fontana[1], BSc, MSc; Paul Aylin[2], MBChB

[1]Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[2]School of Public Health, Imperial College London, London, United Kingdom

[3]Mental Health Innovations, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Lindsay H Dewa, BA, MSc, PhD
Institute of Global Health Innovation
Imperial College London
Reynolds Building, 3rd Floor
London, W6 8RP
United Kingdom
Phone: 44 020 7594 0815
Email: l.dewa@imperial.ac.uk

Abstract

Background: Disrupted social connections may negatively affect youth mental health. In contrast, sustained quality social connections (QSCs) can improve mental health outcomes. However, few studies have examined how these quality connections affect depression and anxiety outcomes within digital interventions, and conceptualization is limited.

Objective: The aim of this study is to conceptualize, appraise, and synthesize evidence on QSC within digital interventions (D-QSC) and the impact on depression and anxiety outcomes for young people aged 14-24 years.

Methods: A systematic scoping review and meta-analysis was conducted using the Joanna Briggs Institute methodological frameworks and guided by experts with lived experience. Reporting was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). The MEDLINE, Embase, PsycINFO, and CINAHL databases were searched against a comprehensive combination of key concepts on June 24, 2020. The search concepts included young people, digital intervention, depression, anxiety, and social connection. Google was also searched. A reviewer independently screened abstracts and titles and full text, and 9.99% (388/3882) of these were screened by a second reviewer. A narrative synthesis was used to structure the findings on indicators of D-QSC and mechanisms that facilitate the connection. Indicators of D-QSC from the included studies were synthesized to produce a conceptual framework.

Results: Of the 5715 publications identified, 42 (0.73%) were included. Among the included studies, there were 23,319 participants. Indicators that D-QSC was present varied and included relatedness, having a sense of belonging, and connecting to similar people. However, despite the variation, most of the indicators were associated with improved outcomes for depression and anxiety. Negative interactions, loneliness, and feeling ignored indicated that D-QSC was not present. In 24% (10/42) of the applicable studies, a meta-analysis showed a significant decrease in depression (–25.6%, 95% CI –0.352 to –0.160; $P<.001$) and anxiety (–15.1%, 95% CI –0.251 to –0.051; $P=.003$) after a D-QSC. Digital mechanisms that helped create a quality connection included anonymity, confidentiality, and peer support. In contrast, mechanisms that hindered the connection included disconnection from the real world and inability to see body language. Data synthesis also identified a 5-component conceptual framework of D-QSC that included rapport, identity and commonality, valued interpersonal dynamic, engagement, and responded to and accepted.

Conclusions: D-QSC is an important and underconsidered component for youth depression and anxiety outcomes. Researchers and developers should consider targeting improved QSC between clinicians and young people within digital interventions for

depression. Future research should build on our framework to further examine relationships among individual attributes of QSC, various digital interventions, and different populations.

(J Med Internet Res 2021;23(12):e26584) doi:10.2196/26584

KEYWORDS

mental health; digital interventions; young people; quality social connection; depression; anxiety; systematic review; meta-analysis; patient and public involvement; mobile phone

Introduction

Background

Enforced lockdowns and physical distancing measures introduced to slow the COVID-19 infection rate have resulted in disrupted face-to-face connections. Ordinarily, a lack of meaningful social connections through social isolation is associated with poor health outcomes such as sleep problems, loneliness, depression, and anxiety, leading in some cases to suicide. Young people are particularly vulnerable to mental health difficulties such as depression and anxiety because onset usually occurs before the age of 24 years [1], and they are often comorbid globally [2]. Although disrupted social connections and loneliness can have a negative effect on mental health [3,4], feeling socially connected is one of the strongest protective factors for depression [5] and can decrease symptoms of anxiety [6].

Social connection as a concept is multifaceted. It can be described as the quantity of connections, the opposite of loneliness, or as having social support. In our context, social connection is the perceived value of attributes of a meaningful interaction among 2 or more people or a *quality* social connection (QSC). Such valued attributes can include, for example, feeling listened to, understood, and a sense of belonging. Similarly, a cooperative relationship between client and therapist, comprising a close bond, shared goals, and tasks, is defined as a *therapeutic alliance* in face-to-face therapy [7]. A therapeutic alliance has been shown to significantly modulate treatment outcomes [8], including in digital settings [9]. Similarly, *social prescribing* to improve social connection has decreased loneliness and improved health outcomes [10]. However, studies have only subjectively demonstrated the value of strong social networks and social relationships for both physical and mental health [11] and longevity [12]. This suggests a need for well-defined indicators of social connection to enable objective measurement of these effects.

The COVID-19 pandemic has accelerated a rapid shift to digital provision of formal and informal mental health support [13]. Indeed, mental health care is often seen as the best candidate for a *digital revolution* because prevention and treatment, including talking therapies, are amenable to delivery over screens and remotely [14]. Social media, video consultations, texting, and virtual reality are interventions that can enable social connections [13]. They represent an important intervention for young people with mental health difficulties to strengthen new and existing relationships and facilitate peer-to-peer and formal mental health support [15]. However, digital interventions such as social media use are also associated with negative consequences such as cyberbullying, viewing harmful content, and a greater sense of isolation [16]. This contradiction requires further investigation to identify the ways in which digital interventions may help or hinder QSC.

Young people are the most digitally fluent and most in need of mental health support. However, QSC within digital interventions (D-QSC) has received little attention in relation to outcomes for depression and anxiety in young people. A recent review produced a conceptual framework for *social connectedness* but positioned it as a solution to loneliness and not as an *active ingredient* (best bet) for the prevention and treatment of depression and anxiety [17]. It also did not consider digital interventions or young people. A systematic review is needed to help produce a conceptual framework for indicators of D-QSC that can be applied to examine its influence on depression and anxiety outcomes across contexts. Our study aims are to (1) identify indicators of D-QSC and their ability to improve or worsen outcomes for depression and anxiety in young people across contexts, (2) identify digital intervention mechanisms that facilitate QSC, and (3) produce a conceptual framework for indicators of D-QSC.

Research Questions

The research questions are:

1. What indicates the presence of QSC in nondigital and digital interventions?
2. How does D-QSC improve or worsen outcomes for depression and anxiety in young people?
3. What digital intervention mechanisms facilitate QSC when preventing or treating depression and anxiety in young people?
4. Whom does D-QSC help or hinder across different contexts, user preferences, and levels of engagement?

Methods

Design

Overview

This systematic scoping review was conducted using the Joanna Briggs Institute methodological framework for scoping reviews. Reporting was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1) guidelines to ensure clear structure, reproducibility, and rigor.

Defining Objectives and Questions and Developing Inclusion Criteria

The research questions were considered, refined, and then finalized with all team members. The Population, Intervention, Comparison, Outcomes, and Study Design Tool was used to

produce our inclusion and exclusion criteria (Table 1). Notably, *young people* (population) as a definition is heterogeneous. However, we have chosen the age group of 14-24 years because it captures key points of vulnerability to developing anxiety and depression between midadolescence and emerging adulthood.

Table 1. Selection criteria.

Category	Inclusion criteria	Exclusion criteria
Population	• Young people aged 14-24 years • Young people aged 14-24 years and additionally 1 year either side of this range (eg, young people aged 13-16 years would be included, whereas those aged 16-26 years would be excluded)	• Nonhuman subjects • Adults aged ≥25 years if unable to easily separate results from those of younger group
Intervention	• Explores QSC[a] (ie, mentions relevant attributes such as empathy, feeling listened to, understood by another person) • Use of a digital intervention, software, or internet-delivered services (eg, smartphone app, virtual reality packages, internet-based treatment, and chat room)	• No mention of QSC (eg, focuses only on quantity of connections) • No mention of digital intervention (eg, based on a face-to-face situation only)
Comparator	• N/A[b]	• N/A
Outcome	• Scope of depression and anxiety spanned all forms, including major, bipolar, psychotic, perinatal, postpartum, PMDD[c], and manic depression, as well as social, generalized, OCD[d], panic, PTSD[e], and anxiety disorders • Influence on existing symptoms of depression or anxiety (eg, mood and self-esteem through self-report questionnaire or clinical interview) • Prevention of onset of depression or anxiety (eg, measuring mental well-being through self-report questionnaire or clinical interview)	• No mention of depression or anxiety • No mention of the influence on existing symptoms of depression or anxiety • No mention of the influence on depression or anxiety prevention
Study design	• All study designs	• N/A
Dates	• From earliest date to June 24, 2020	• Outside date remit

[a]QSC: quality social connection.

[b]N/A: not applicable.

[c]PMDD: premenstrual dysphoric disorder.

[d]OCD: obsessive-compulsive disorder.

[e]PTSD: posttraumatic stress disorder.

Searching for the Evidence

The MEDLINE, Embase, PsycINFO, and CINAHL databases were searched on June 24, 2020. The search strategy was developed and verified by 3 team members (LD, EL, and HA) and an institutional librarian and tailored to each database (Multimedia Appendix 2 [18-59]). In all, 4 facets made up the strategy, including young people (eg, *youth* and *teens*), social connection (*social connect** and *sociali?ation*), digital intervention (eg, *online* and *digital*), and depression and anxiety (*depress** and *anx**). The World Health Organization International Clinical Trials Registry Platform, ClinicalTrials.gov, and the *Journal of Medical Internet Research* were searched on July 14, 2020. The first 100 Google search hits were also systematically searched by 2 reviewers (LR and EBH) using key words across the 4 facets (eg, young people, social connect*, anxiety and depression, and digital) as a further check (Multimedia Appendix 2). The included papers' reference lists were also reviewed and added to the search if appropriate.

Selecting the Evidence

Titles and abstracts were independently screened by 1 reviewer (LR) and excluded if they did not match the selection criteria (Table 1). Studies that met the inclusion criteria were retrieved in full by the primary reviewer (LR) and reassessed against the selection criteria in detail. A second reviewer (EBH) independently screened a random 9.99% (388/3882) of the titles, abstracts, and full-text manuscripts to ensure reliability in study selection. A predefined interreliability agreement (≥0.70) was agreed upon and calculated. Another random 9.99% (388/3882) would have been screened until agreement was achieved. Disagreements were resolved with a third reviewer (LD).

Extracting and Charting the Evidence

The data-charting process documented indicators of QSC, prevention and treatment categorization, digital intervention mechanisms that facilitate QSC, and participant characteristics. An initial 20% (8/42) of the studies were extracted independently by 2 reviewers (LR and EBH) and reviewed to ensure accuracy before 1 reviewer (LR) continued with the remaining extraction. All included studies were also appraised using the Hawker checklist [60], which is designed specifically for cross-comparison across heterogeneous designs (quantitative, qualitative, and mixed methods). A total of 9 domains were appraised: (1) abstract and title, (2) introduction and aims, (3) methods and data, (4) sampling, (5) data analysis, (6) ethics and bias, (7) results, (8) transferability and generalizability, and (9) implications and usefulness. Quality scores were assigned to

each domain, from 1 point (very poor) to 4 points (good), summed and assigned as high quality (30-36 points), medium quality (24-29 points), or low quality (9-23 points).

Analysis of the Evidence, Presentation of the Results, and Summarizing the Evidence

Meta-analyses were performed where appropriate to examine the effect of D-QSC on outcomes. Overall and specific categories of depression, anxiety, and well-being outcomes were analyzed by calculating the ratio of means within each study. We substituted median for mean in studies where only the median was reported. The inverse-variance, random-effects model of DerSimonian and Laird [61] was used for analysis of both continuous and categorical variables in Stata software (version 15; StataCorp) [62]. The I^2 statistic was used to estimate the degree of heterogeneity among studies, where larger values indicated increasing heterogeneity. The scoping nature of the review meant that a narrative approach was appropriate. All indicators of the development and presence of D-QSC were first collated and synthesized using a deductive approach. The initial relationship between these indicators and the outcomes was explored. Potential themes were identified, discussed, and agreed upon by 3 reviewers (LD, LR, and EL).

To produce a conceptual framework for indicators of D-QSC there were 4 main stages. At the first stage, all indicators identified in the literature synthesis or by experts with lived experience (see the *Patient and Public Involvement* section) were added as cards in Miro (ie, participative visual platform). Indicators that directly described *social connection* (eg, social connectedness) were repeated and those that were not an attribute of D-QSC were excluded. Second, the remaining indicators were either grouped with similar indicators or stood alone. Third, the indicators were then mapped onto a preexisting framework of the components of social connectedness in mental disorders (closeness, identity and common bond, valued relationships, involvement, and cared for and accepted [CIVIC] framework) [17]. Indicators that did not map onto the preexisting framework were kept together and merged under a new component name. This resulted in preliminary components of D-QSC. Finally, young experts with lived experience critically reflected on the preliminary framework and answered a series of questions at a web-based meeting and through email. For example, “Is any indicator missing?” and “Do the indicators link together well or should they be moved?” This discussion was unstructured to allow independent and novel thought. As a result, changes were made to either component or indicator wording and indicators were added or merged. All team members and the young experts agreed on the final conceptual framework for D-QSC.

Patient and Public Involvement

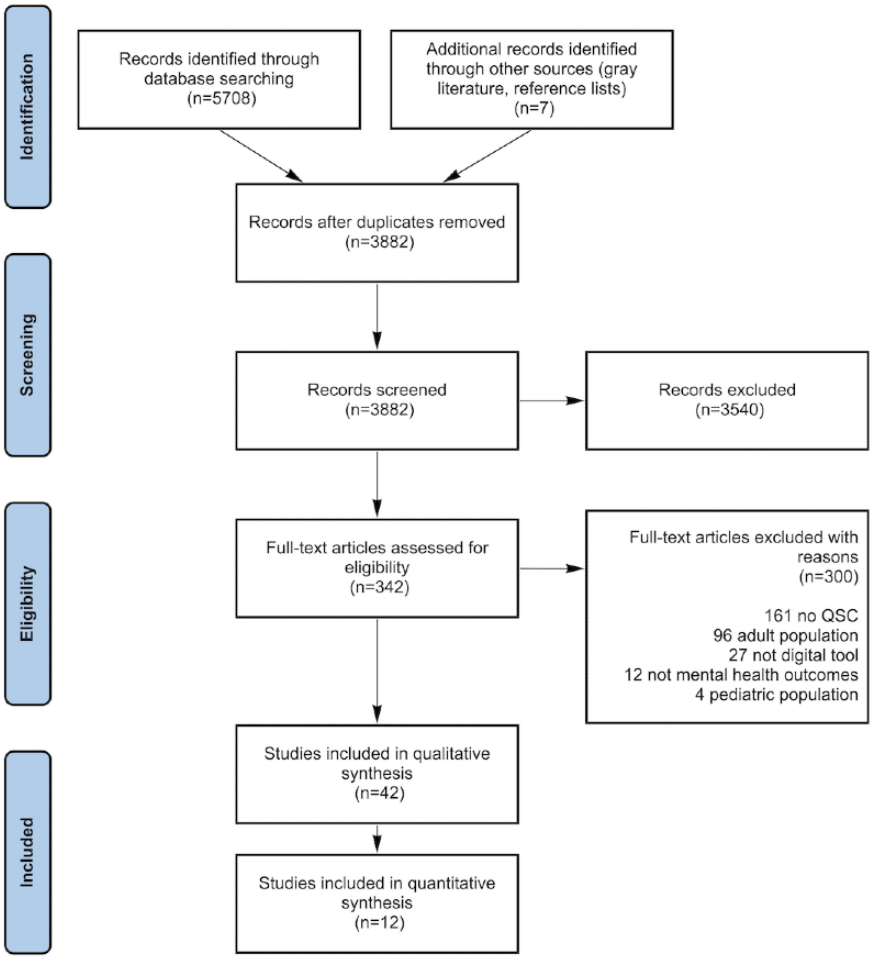
We advertised for young people aged 14-24 years with experience of depression or anxiety and digital interventions for mental health to work on a review about social connection in the digital world through The McPin Foundation newsletter, email distribution lists, Twitter, and Instagram. A total of 9 people applied using a simple form, and all joined the Young Persons Advisory Group (YPAG). They represented different genders, ethnicities, ages, and UK locations. We held an initial web-based workshop to help define QSC and D-QSC, inform search terms, and review the protocol. At this stage, we approached the Lancet Commission for Global Mental Health Young Leaders and experts with professional experience (ie, delivery of digital interventions) to ensure a diverse range of experiences, cultural contexts (ie, low- and high-resource settings), and experience of youth interventions for depression or anxiety. We had separate discussions with each group on the web (eg, Zoom). Subsequent changes were made to our definition of D-QSC, selection criteria, and protocol. On the basis of definitions of quality [63] and therapeutic alliance [7] and input from team members (EL, LD, and LR) and experts, QSC was then operationally defined as *theperceived value of the attributes of an interaction between two or more people*. Key attributes (ie, indicators) of D-QSC that made up the definition were logged across the 3 expert groups and amalgamated with the literature indicators as described previously. Others were changed (eg, changed to plain English) or merged after the YPAG and the Lancet Commission for Global Mental Health Young Leaders were shown the findings and conceptualization framework. A YPAG member (EBH) also screened, extracted, and quality-assessed literature. All were given appropriate support and paid in line with guidance [64].

Results

Overview

A total of 5715 records were identified (Figure 1). Of the 5715 articles, 1833 (32.07%) were duplicates and were removed. Substantial agreements were achieved in the screening of the random 9.99% (388/3882) abstracts and titles as well as full-text subsamples (K=0.80 and K=0.70, respectively). Papers were then excluded if they did not match the selection criteria; of the 3882 publications remaining after duplicates were removed, 42 (1.08%) were included (Figure 1). Of these 42 studies, 28 (67%) were of high quality, 13 (31%) were of medium quality, and 1 (2%) was of low quality. High-quality studies largely demonstrated good explanation of aims, methods, and sampling to enable replicability. Medium-quality studies included most of the good study criteria but were lacking in some areas, which reduced their scores. The low-quality study did not provide enough detail across most domains (eg, it did not report ethical considerations, the results were unclear, and the methods were not replicable).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. QSC: quality social connection.



Study and Participant Characteristics

Of the 42 included studies, 25 (60%) were quantitative [18-42], 11 (26%) were qualitative [43-53], and 6 (14%) were mixed methods studies [54-59] (Table 2; Multimedia Appendix 3 [18-59]). The studies mainly used uncontrolled or cross-sectional designs and had questionnaires as their main data collection method (22/42, 52%). All studies took place in high-income countries: the United States [23-28,36,38,40,42,44,47,49,50,53], Australia [22,29-31,43,46,55-58], Ireland [18,45,54], Israel [19,32,48], Taiwan [20,35], Sweden [21,41], the Netherlands [33], Turkey [39], Austria [34], Belgium [37], Canada [59], Cyprus [52], and the United Kingdom [51]. There were 23,319 participants in total (12,825/23,319, 55%, were women; of the 42 studies, 2 (5%) did not report participant gender). Of the 42 studies, 21 (50%) focused on both prevention and treatment [18,20,21,23-29,36-39,41-44,48,50,57], 12 (29%) focused on treatment [22,30,31,45,49,51-56,58], and 9 (21%) focused on prevention [19,32-35,40,46,47,59]. Digital mental health–related intervention types included mental health social networking tools, smartphone apps, self-help cognitive behavioral therapy, telepsychiatry, one-to-one peer mentor support, video gaming, avatars, and internet use for mental health support. Nonspecific informal digital interventions included general social networking and social media (eg, Facebook, Twitter, Tumblr, Snapchat, and Reddit) and general internet use and web browsing. Intervention duration was reported in 48% (20/42) of the studies and ranged from 8 weeks to 1 year.

Table 2. Data extraction and quality assessment of included studies (N=42).

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Alvarez-Jimenez et al [58], 2013, Australia, high quality	Quantitative and qualitative, uncontrolled single-group, observation, questionnaire, and semistructured interview	Setting: early psychosis prevention and intervention center; sample: 20 patients (50% female; aged 15-25 years; 45% Anglo-Australian, 25% Asian, 10% biracial, and 5% African); presenting condition: first episode psychosis	Peer-to-peer web-based social networking, individually tailored web-based psychosocial interventions, and expert moderation: HORYZONS	Outcomes: depression and anxiety reduced; measures: BPRS[a], CDRS[b], and BAI[c]
Alvarez-Jimenez et al [30], 2018, Australia, medium quality	Quantitative, uncontrolled single-group, observation, and semistructured interview	Setting: PACE[d] clinic for ultrahigh-risk psychosis; sample: 14 patients (79% female; aged 15-25 years; ethnicity unknown; all Australia-born); presenting condition: ultra-high risk for psychosis	Web-based social networking, peer-to-peer and professional moderation: MOMENTUM	Outcomes: depression reduced and psychological well-being improved; measures: SWLS[e], MADRS[f], and PSSS[g]
Bailey et al [22], 2020, Australia, high quality	Quantitative, uncontrolled single-group pre- and posttest, observation, and semistructured interview	Setting: tertiary-level mental health service; sample: 20 patients (55% female; aged 16-25 years; ethnicity unknown; country of birth: 75% Australia, 20% Asia, and 5% United Kingdom); presenting condition: suicidal ideation	Enhanced web-based social networking intervention: Affinity	Outcome: depression reduced; measure: PHQ-9[h]
Bhuvaneswar and Gutheil [49], 2008, United States, high quality	Qualitative, retrospective case study, and observation	Setting: psychodynamic psychotherapy clinic; sample: 1 patient (female; aged 17 years; ethnicity unknown); presenting condition: depression	Instant messenger	Outcome: psychological well-being worsened; measure: self-report
Blackwell et al [24], 2012, United States, high quality	Quantitative, randomized controlled trial, and questionnaire	Setting: general; sample: 100 adolescents (62% female; mean age 15.69 years, SD 2.91 years; 57% White, 16% Hispanic, 9% African American, and 18% ethnicity unknown); presenting condition: cystic fibrosis	Web-based social networking peer support program: CFfone.com	Outcomes: Depression and anxiety reduced; measure: HADS[i]
Campbell et al [55], 2019, Australia, medium quality	Qualitative and quantitative, participatory action research design, observation, and questionnaire	Setting: Kids Helpline family discord service; sample: 105 callers to helpline (82% female; aged 13-25 years; ethnicity unknown); presenting condition: mild to moderate depression or anxiety (not high risk)	Social networking site for peer-to-peer and counsellor-to-peer group support	Outcomes: depression and anxiety—data quality too low to assess; measures: CES-D[j] and RCMAS[k]
Canady [25], 2018, United States, high quality	Quantitative, cross-sectional study, questionnaire, and interview	Setting: general; sample: 1300 adolescents (gender unknown; aged 14-22 years; ethnicity unknown); presenting condition: none in particular	Web-based health information and digital health tools in general, including peer-to-peer health exchange networks	Outcomes: Depression and anxiety reduced; measures: PHQ-9 and self-report
Chyzzy et al [59], 2020, Canada, high quality	Qualitative and quantitative, uncontrolled single-group design, questionnaire, and semistructured interview	Setting: MPPS[l] intervention group; sample: 21 mothers (100% female; aged 17-24 years, mean age 21.3, SD 1.8, years; ethnicity unknown; country of birth: 66.7% Canada), presenting condition: generally healthy, 14.3% with prior history of depression	Individualized peer mentor support through telephone call and SMS text messaging: MPPS intervention	Outcomes: depression and anxiety reduced; measure: self-report
Clarke [45], 2018, Ireland, high quality	Qualitative, retrospective case study, and observation	Setting: clinical; sample: 1 patient (male; aged 16 years; ethnicity unknown); presenting condition: Asperger syndrome with comorbid depression	Telepsychiatry	Outcome: depression treatment engagement improved; measure: observation

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Colder Carras et al [28], 2017, United States, medium quality	Quantitative, cross-sectional study, and questionnaire	Setting: 30 US schools; sample: 9733 students (51% female; aged 13-16 years, average age 14.1 years; 82.1% Dutch); presenting condition: none in particular	Web-based video gaming	Outcomes: depression and social anxiety reduced for *social engaged gamers* compared with *problematic*, *at-risk*, or *extensive* gamers; measures: depressive mood list and SASC-R[m]
Cole et al [36], 2017, United States, —[n]	Quantitative, uncontrolled single-group design, and questionnaire	Setting: private university; sample: 231 undergraduate students (72% female; average age 19.28, SD 1.15, years; 67.1% White, 23.4% Asian American, 10.4% African American, 5.2% Hispanic or Latino, and 0.4% Other); presenting condition: none in particular	Web-based social networks in general	Outcomes: depression worsened; measures: DASS[o], CTI[p], and BDI-II[q]
Dhesi [51], 2019, United Kingdom, high quality	Qualitative, cross-sectional, and web-based semistructured interviews	Setting: Kooth digital mental health care service; sample: 13 Kooth users (69% female; aged 14-18 years; 69.2% White British, 15.4% White and Asian, and 15.4% Other); presenting condition: none in particular	Web-based counseling (text)	Outcomes: anxiety reduced; measure: thematic analysis of interviews
Dolev-Cohen and Barak [48], 2013, Israel, high quality	Qualitative, case-control design, questionnaire, textual analysis, and observation	Setting: general; sample: 150 instant messaging users (63% female; aged 14-18 years; ethnicity unknown); presenting condition: distressed vs nondistressed groups of participants	Regular use of instant messaging	Outcome: psychological well-being improved; measure: PANAS[r]
Ellis et al [56], 2011, Australia, —	Qualitative and quantitative, comparative randomized controlled trial, and questionnaire	Setting: university students not receiving mental health treatment; sample: 39 students (77% female; aged 18-25 years, mean age 19.67, SD 1.66, years; ethnicity unknown); presenting condition: anxiety or depression but none severe	Web-based cognitive behavior therapy self-help program (MoodGYM) compared with web-based support group (MoodGarden)	Outcomes: depression and anxiety reduced; measures: DASS and ATQ[s]
Feinstein et al [26], 2012, United States, high quality	Quantitative, short-term prospective cohort study, and questionnaire	Setting: undergraduate university students; sample: 301 students (62% female; mean age 19.44, SD 2.05, years; 41% Asian or Pacific Islander, 41% White, 6% Latino, 6% African American, and 6% Other); presenting condition: some participants had raised depression, anxiety, or social anxiety at baseline	Social networking in general	Outcome: depression resulted in poor-quality social connections, which in turn worsened depression and anxiety; measures: DASS and BFNE[t]
Felnhofer et al [34], 2018, Austria, —	Quantitative, randomized controlled trial, and questionnaire	Setting: public university; sample: 95 students (87% female; mean age 23.34, SD 2.727, years; ethnicity unknown); presenting condition: none in particular	Avatars (virtual entities controlled by another human being) and agents (virtual entities controlled by a computer)	Outcome: social interaction anxiety unchanged; measure: SIAS[u]
Frison and Eggermont [37], 2016, Belgium, medium quality	Quantitative, uncontrolled cross-sectional, and questionnaire	Setting: 18 randomly selected high schools in Flanders, Belgium; sample: 910 students with Facebook account (52% female; average age 15.44, SD 1.71, years; ethnicity unknown; country of birth: 96.1% Belgium, 1.8% Europe, and 2.1% non-European country); presenting condition: none in particular	Facebook	Outcome: depression reduced; measure: CES-DC[v]

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Garrido et al [43], 2019, Australia, medium quality	Qualitative and focus groups	Setting: high schools and universities in Western Australia; sample: 23 students (65% female; aged 13-25 years; ethnicity unknown); presenting condition: DASS score <15 (severely depressed excluded)	A total of 6 currently available smartphone apps for mental health (Mood Mission, Music eScape, Pacifica, Mindshift, Headspace, and What's Up)	Outcome: helpful and unhelpful aspects of smartphone apps for mental health; measure: thematic analysis of focus group content
Horgan et al [54], 2013, Ireland, medium quality	Qualitative and quantitative, pre- and posttest and qualitative descriptive designs, extraction of posts from website, and questionnaire for CES-D scores	Setting: University of Cork; sample: 118 students (36% female; aged 18-24 years; 98.3% White and 1.7% Asian or Asian Irish); presenting condition: depression	Depression support website with peer support forum	Outcome: depression reduced; measure: CES-D
Horgan and Sweeney [18], 2010, Ireland, medium quality	Quantitative, descriptive study, and questionnaire	Setting: university; sample: 922 students (62% female; aged 18-24 years; ethnicity unknown); presenting condition: none in particular	Internet use for mental health support	Outcome: reasons for use of internet-based mental health support; measure: self-developed questionnaire
Lim et al [57], 2019, Australia, high quality	Qualitative and quantitative, descriptive design, pre- and posttest questionnaires, mood tracker, and semistructured interview	Setting: local youth health service (participants with social anxiety disorder) and Australian university (participants without social anxiety disorder); sample: 20 participants (45% female; aged 18-23 years; 91% White and 9% multiracial or other); presenting disorder: with or without social anxiety disorder	+Connect, a digital smartphone app with video material	Outcomes: depression and anxiety reduced; measures: CES-D and SIAS
Liu and Yu [35], 2013, Taiwan, medium quality	Quantitative, cross-sectional study, and questionnaire	Setting: college; sample: 330 Facebook-using students (63% female; aged 18-23 years; ethnicity unknown); presenting condition: none in particular	Facebook	Outcome: psychological well-being improved; measure: Ryff scales of psychological well-being
McCloskey et al [23], 2015, United States, medium quality	Quantitative, uncontrolled single-group design, and questionnaire	Setting: university; sample: 633 undergraduate students with Facebook page (70% female; aged ≥18 years, median age 21 years; 64.8% White); presenting condition: none in particular; participants on average had mild levels of depression at baseline	Facebook	Outcome: depression reduced; measure: PHQ-9
Mikami [38], 2010, United States, high quality	Quantitative, longitudinal, observation, and questionnaire	Setting: public middle school; sample: 92 social networking site users (58% female; mean age 20.92, SD 1.11, years; 58% White, 29% African American, and 13% Other or Mixed); presenting condition: none in particular	Web-based social networking	Outcome: depression—no outcomes reported; measure: CDI[w]
Ozcan and Buzlu [39], 2007, Turkey, high quality	Quantitative, uncontrolled single-group design, and questionnaire	Setting: university; sample: 730 undergraduate students who use the internet (53% female; mean age 20.84, SD 1.95, years; ethnicity unknown); presenting condition: none in particular	Internet use in general	Outcome: depression reduced; measure: BDI
Poppelaars [33], 2018, The Netherlands, medium quality	Quantitative, randomized controlled trial, and questionnaire	Setting: university; sample: 146 undergraduate students who play video games (71% female; mean age 20.2, SD 1.74, years; ethnicity unknown; nationality: 76% Dutch, 23% German, and 1% Other); presenting condition: none in particular; some with higher depressive symptoms at outset	Video game that included cooperation with other players and with mental health messaging vs video game without mental health messaging	Outcome: psychological well-being improved, with larger improvement for those higher in depressive symptoms; measures: BDI-II, SAM[x], and International PANAS short form

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Radovic [44], 2017, United States, United States, high quality	Qualitative, randomized controlled trial, semistructured interviews, think aloud, advisory boards, and focus groups	Setting: academic adolescent medicine clinic and specialty psychiatry clinic; sample: 23 patients (78% female; aged 13-20 years, mean age 16, SD 2.3, years); presenting condition: depression	Social media website for depressed adolescents	Outcome: adolescent-informed design of social media website for depression; measure: thematic analysis from semistructured interviews
Radovic [53], 2017, United States, medium quality	Qualitative, uncontrolled cross-sectional study, and semistructured interview	Setting: academic adolescent medicine clinic and specialty psychiatry clinic; sample: 23 patients (78% female; aged 13-20 years, mean age 16, SD 2.3, years; 87% White); presenting condition: depression	Social media	Outcomes: depressive symptoms either made participants reach for social media as a distraction or avoid it to avoid bringing down others. Psychological well-being improved; measure: thematic analysis from semistructured interviews
Rice et al [29], 2018, Australia, medium quality	Quantitative, uncontrolled single-group pilot, structured clinical interview, and questionnaire	Setting: mental health clinic; sample: 42 patients (50% female; aged 15-25 years, mean age 18.5, SD 2.1, years; ethnicity unknown; country of birth: 95.2% Australia); presenting condition: previous depression sufferers	Novel, moderated web-based social therapy intervention: Rebound	Outcomes: depression reduced and anxiety unchanged; measures: MADRS and DASS
Rice et al [31], 2020, Australia, high quality	Quantitative, single-group uncontrolled pre-post design, and questionnaire	Setting: 4 Headspace early intervention centers in northwestern Melbourne; sample: 89 patients (47% female; aged 14-25 years; ethnicity unknown); presenting condition: social anxiety	Social networking platform for socially anxious young people (Entourage): a *wall* function allows posting and commenting publicly	Outcomes: depression and social anxiety reduced and psychological well-being improved; measures: PHQ-9, MDRS-22[y], LSAS[z], BFNE, SIAS, and SWEMWBS[aa]
Santesteban-Echarri et al [46], 2017, Australia, medium quality	Qualitative, uncontrolled single-group pilot, semistructured interview, and focus group data	Setting: mental health clinic; sample: 42 patients (50% female; aged 15-25 years, mean age 18.5, SD 2.1, years; ethnicity unknown; country of birth: 95.2% Australia); presenting condition: previous depression sufferers	Novel, moderated web-based social therapy intervention: Rebound	Outcome: efficacy and usability evaluation of web-based social therapy intervention; measure: thematic analysis from semistructured interviews
Saulsberry et al [40], 2013, United States, medium quality	Quantitative, randomized controlled trial, and telephone interview	Setting: 12 primary care sites across southern and midwestern United States; sample: 58 patients (57% female; mean age 17.26, SD 1.85, years; 61% White, 24% Black, 6% Asian, 5% Hispanic, and 4% Other); presenting condition: depression	Primary care provider motivational interview+CATCH-IT internet program vs primary care provider brief advice+CATCH-IT internet program	Outcome: depression reduced; measures: CES-D-10, DSM-IV-TR[ab], and PHQ-A[ac]
Selkie et al [47], 2020, United States, high quality	Qualitative, uncontrolled single-group design, and semistructured interviews	Setting: pediatric gender clinic; sample: 25 transgender adolescents with social media profile (44% trans-feminine; aged 15-18 years, mean age 16 years; 80% White non-Hispanic, 4% African American, 8% American Indian, and 8% Asian); presenting condition: none in particular	Social media platforms, including YouTube, Instagram, Facebook, Twitter, and Tumblr	Outcomes: positive and negative outcomes of using social media for mental health support; measure: —
Sharabi and Margalit [32], 2011, Israel, medium quality	Quantitative, cross-sectional crossover, and questionnaire	Setting: middle to high socioeconomic families vs those who failed in school (mostly from low socioeconomic families); sample: 716 students (48% female; aged 16-18 years; ethnicity unknown); presenting condition: with or without learning disabilities	Internet communication	Outcomes: psychological well-being negatively correlated with loneliness. Loneliness reduced by internet communication with people known offline; measure: Hebrew adaptation of Mood Scale

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Sharabi and Margalit [19], 2011, Israel, medium quality	Quantitative and cross-sectional case-control	Setting: 3 high schools in urban Israel; sample: 887 students grades 10-12 (50% female; aged 16-18 years; ethnicity unknown); presenting condition: with (n=213) or without (n=674) learning disabilities	Internet communication	Outcome: psychological well-being reduced; measure: Hebrew adaptation of Affect Scale
Siriaraya et al [52], 2011, Cyprus, medium quality	Qualitative, cross-sectional study, and content analysis	Setting: general; sample: 400 messages from teenagers using web-based discussion forum (gender unknown; age range unknown; ethnicity unknown); presenting condition: none in particular	Web-based anonymous discussion forum	Outcome: level of support provided among adolescents; measure: Content analysis of web-based forum messages
Stockdale and Coyne [27], 2020, United States, high quality	Quantitative, longitudinal, and questionnaire	Setting: longitudinal study of intrafamily life participants; sample: 385 participants who use smartphones (53% female; aged 17-19 years; 70% European-American, 10% African American, 12% Multiethnic, 5% Asian American, and 2% Other); presenting condition: none in particular	Social media use	Outcomes: depression unchanged and anxiety worsened; measures: CES-DC and SCAS[ad]
van Rensburg et al [50], 2015, United States, high quality	Qualitative, uncontrolled single-group design, and semistructured interviews	Setting: Yale Psychiatric Hospital Intensive Outpatient Program; sample: 20 patients (75% female; aged 14-19 years; 80% White, 15% Hispanic, and 5% Mixed); presenting condition: combination of ADHD[ae], mood disorder NOS[af], MDD[ag], anxiety, PTSD[ah], psychosis, and ODD[ai]	Social media for patient-provider interactions	Outcomes: positive (including safety) and negative (including anxiety) outcomes of patient-provider interactions through social media; measure: thematic analysis of semistructured interviews
van Zalk et al [41], 2011, Sweden, high quality	Quantitative, uncontrolled single-arm longitudinal study, and questionnaire	Setting: university in Utrecht; sample: 197 psychology freshmen (78% female; mean age 18.9, SD 1.6, years; ethnicity unknown; 92% Dutch origin); presenting condition: none in particular	Web-based chatting with friends through web-based social networking site	Outcome: depression unchanged; measure: BDI Dutch short version
Van Zalk and Tillfors [21], 2017, Sweden, high quality	Quantitative, longitudinal study, and questionnaire	Setting: Swedish school; sample: 526 students from grades 7-9 (68% female; aged 13-15 years; ethnicity unknown; 12.1% first-generation immigrants); presenting condition: none in particular	Web-based chatting with friends through web-based social networking site	Outcome: Reduced depression among those with higher, but not lower, social anxiety; measures: CES-D and SPSQ-C[aj]
Wright et al [42], 2013, United States, medium quality	Quantitative, cross-sectional observational study, and questionnaire	Setting: undergraduate university; sample: 361 students who use Facebook (53% female; mean age 20.26, SD 2.72, years; 77% White, 8.6% Native American, 4.4% Latino, 3.6% Asian American, 3.3% African American, and 3.3% Other); presenting condition: none in particular	Facebook use	Outcome: depression reduced; measure: CES-D

Author, year, country, quality	Study design	Setting and participants	Digital intervention	Outcomes and measures
Yeh et al [20], 2008, Taiwan, medium quality	Quantitative, cross-sectional, and questionnaire	Setting: project of mental health survey; sample: 3477 college students (55% female; mean age 22.45, SD 1.56, years; ethnicity unknown); presenting condition: none in particular	Social support on the web	Outcome: depression worsened by higher web-based and lower actual social support; measure: Ko Depression Inventory

[a]BPRS: Brief Psychiatric Rating Scale.
[b]CDRS: Children's Depression Rating Scale.
[c]BAI: Beck Anxiety Inventory.
[d]PACE: Personal Assessment and Crisis Evaluation.
[e]SWLS: Satisfaction With Life Scale.
[f]MADRS: Montgomery–Åsberg Depression Rating Scale.
[g]PSSS: Perceived Social Support Scale.
[h]PHQ-9: Patient Health Questionnaire Depression Scale.
[i]HADS: Hospital Anxiety and Depression Scale.
[j]CES-D: Center for Epidemiological Studies Depression Scale.
[k]RCMAS: Revised Children's Manifest Anxiety Scale.
[l]MPPS: Mothers' Perceptions of Mobile Phone–Based Peer Support.
[m]SASC-R: Social Anxiety Scale for Children-Revised.
[n]Not available.
[o]DASS: Depression Anxiety Stress Scales.
[p]CTI: Cognitive Triad Inventory.
[q]BDI-II: Beck Depression Inventory II.
[r]PANAS: Positive and Negative Affect Scale.
[s]ATQ: Automatic Thoughts Questionnaire.
[t]BFNE: Brief Fear of Negative Evaluation.
[u]SIAS: Social Interaction Anxiety Scale.
[v]CES-DC: Center for Epidemiological Studies Depression Scale for Children.
[w]CDI: Children's Depression Inventory.
[x]SAM: Self-Assessment Manikin.
[y]MDRS-22: Male Depression Risk Scale.
[z]LSAS: Liebowitz Social Anxiety Scale.
[aa]SWEMWBS: Short Warwick–Edinburgh Mental Well-being Scale.
[ab]DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision.
[ac]PHQ-A: Patient Health Questionnaire-9 modified for Adolescents.
[ad]SCAS: Spence Children's Anxiety Scale.
[ae]ADHD: attention-deficit/hyperactivity disorder.
[af]NOS: not otherwise specified.
[ag]MDD: major depressive disorder.
[ah]PTSD: posttraumatic stress disorder.
[ai]ODD: oppositional defiant disorder.
[aj]SPSQ-C: Social Phobia Screening Questionnaire for Children and Adolescents.

Indicators That QSC Is Present in Digital Interventions

Indicators and measures used to quantitatively assess D-QSC presence were heterogeneous (Tables 3 and 4). The most common indicator for D-QSC was social support (14/42, 33%; Table 3). Among the 31 quantitative studies assessed, there were 20 different standardized questionnaires used to measure QSC (Tables 3 and 4), with only 4 (13%) studies using the same measure (Multidimensional Scale of Perceived Social Support). Nonstandardized questionnaires were also used in some studies, including single questions (eg, "I hope to gain support through meeting people going through similar experiences, Y/N"). The remaining indicators of QSC were identified from qualitative analysis within 17 studies [43-59].

Table 3. Indicators of the presence of quality social connection within digital interventions in the included studies (N=42).

Indicator	Description and measurement example	Values, n (%)	Improved depression outcomes, n (%)	Improved anxiety outcomes, n (%)
Social support[a]	Level of support received from others. Multidimensional Scale of Perceived Social Support: "There is a special person who is around when I am in need"	14 (33) [20,23,24,30,35-39,41,42,46,53,56]	9 (64) [23,24,30,36,37,39,41,42,56]	2 (14) [24,56]
Social connectedness[a]	A sense of feeling connected to others. Social Connectedness Scale Revised: "I feel understood by the people I know"	10 (24) [22,27-29,31,43,52,56-58]	6 (60) [22,28,29,31,57,58]	5 (50) [28,31,52,57,58]
Relatedness	Bonding through shared experience or understanding. Open-ended survey questions to determine best and worst aspects of intervention	5 (12) [33,43,46,55,56]	1 (20) [56]	1 (20) [56]
Connecting with similar people	Communicating with those who have similar experiences and feelings. Content analysis and thematic coding of qualitative questions	4 (10) [18,25,47,53]	1 (25) [25]	1 (25) [25]
Feeling accepted	Having a sense that people are okay with, and accepting of, oneself. Likert-scale response to statement "I felt that the [forum] moderators accepted me"	3 (7) [29,53,59]	2 (67) [29,59]	1 (33) [59]
Being able to share	Feeling able to disclose one's thoughts and feelings to others. Friendship Quality Questionnaire: "I would tell him or her what upsets me"	4 (10) [21,51,52,54]	2 (50) [21,54]	1 (25) [52]
Feeling normalized	Someone making it clear that what one is feeling is normal. Peer Support Evaluation Inventory subscale item: "Helped me feel that what I was going through was 'normal'"	3 (7) [47,52,59]	1 (33) [59]	2 (67) [52,59]
Feeling close to a peer	A sense of intimacy or connection with another person. Peer Support Evaluation Inventory subscale item: "I felt close to my peer"	2 (5) [57,59]	2 (100) [57,59]	2 (100) [57,59]
Less alone in one's feelings	Knowing that others are experiencing similar feelings. Content analysis and thematic coding of qualitative interview questions	3 (7) [25,54,55]	2 (67) [25,54]	1 (33) [25]
Sense of belonging	Feeling that one is part of a group. Interpersonal Needs Questionnaire: "I don't fit in"	2 (5) [22,31]	2 (100) [22,31]	1 (50) [31]
Emotional connection	A bond created among 2 or more people by sharing feelings. Text-based ethnographic study of instant messaging conversations	2 (5) [48,50]	—[b]	—
Empathy	Understanding and sharing feelings of another person. Networked Minds Measure of Social Presence Empathy subscale: "When the other was happy, I was happy"	2 (5) [27,34]	—	—

Indicator	Description and measurement example	Values, n (%)	Improved depression outcomes, n (%)	Improved anxiety outcomes, n (%)
Feeling you are not a burden	Sense that one is not bothering or troubling others. Interpersonal Needs Questionnaire low score for items such as "These days I think I make things worse for the people in my life"	2 (5) [22,31]	2 (100) [22,31]	1 (50) [31]
Rapport	Trust and understanding established between the provider and patient. Provider-reported from ethnography	1 (2) [45]	—	—
Feeling validated	Having acceptance or approval from others of one's thoughts and feelings. Content analysis and thematic coding of qualitative questions, categorized as *Appraisal support*	1 (2) [47]	—	—
Shared understanding	Another person knowing how one is feeling through their own similar experience. Content analysis and thematic coding of forum posts	1 (2) [54]	1 (100) [54]	—
Trust	Ability to rely on someone. Peer Support Evaluation Inventory: "My peer was trustworthy"	1 (2) [59]	1 (100) [59]	1 (100) [59]

[a]Directly encapsulates the definition of quality social connection.

[b]Not available.

Table 4. Indicators of the absence of quality social connection within digital interventions in the included studies (N=42).

Indicator	Description and measurement example	Values, n (%)	Improved depression outcomes, n (%)	Improved anxiety outcomes, n (%)
Negative interactions	Harm being inflicted through digital interventions, resulting in negative feelings such as loneliness or hurt. Social Networking Survey: "How positive (or negative) are your interactions with people on FB[a] and MS[b]?"	6 (14) [26,43,47,49,51,53]	—[c]	—
Loneliness	A sense of isolation as a result of being disconnected from other people. University of California, Los Angeles, Loneliness Scale: "I lack companionship"	7 (17) [19,30-32,40,47,59]	4 (57) [30,31,40,59] (Reduced loneliness)	2 (29) [31,59] (Reduced loneliness)
Feeling ignored	Not being responded to. Content analysis and thematic coding of semistructured interviews exploring engagement with therapist through social networks and its efficacy	2 (5) [49,51]	—	—

[a]FB: Facebook.

[b]MS: Myspace.

[c]Not available.

Associations Between D-QSC and Outcomes

The relationship between D-QSC indicators and outcomes was mixed. Of the 42 studies, 10 (24%) reported a change in depression symptoms over time after participants experienced a D-QSC, and a pooled analysis demonstrated a significant weighted mean decrease in depression by 25.6% (–0.256, 95% CI –0.352 to –0.160; P<.001), with high heterogeneity (I^2= 90.8%; Figure 2) [20,23,24,26,30,36,37,39,41,42]. Of the 42 studies, 5 (12%) reported change over time in anxiety symptoms; there was also a decrease, but it was smaller (15%; –0.151, 95% CI –0.251 to –0.051; P=.003), with high heterogeneity (I^2= 83.1%; Figure 3) [29,31,56-58].

The indicators of D-QSC associated with improved depression or anxiety symptoms included social support [23,24,30,36,37,39,41,42,56], social connectedness [22,28,29,31,52,57,58], loneliness (reduced) [30,31,40,54,59],

relatedness [56], sense of belonging [22,31], being able to share [21,52,54], less alone in one's feelings [25,54], feeling normalized [52,59], feeling close to peer [57,59], feeling you are not a burden [22,31], feeling accepted [29,59], shared understanding [54], and trust [59] (Table 3). For example, depression outcomes improved after good social support for those abused on the web [36] and for adolescents with high social anxiety [21]. In contrast, negative interactions [20,26], negative experiences of social support [49,51,52], and frequent social media use [27] were associated with worsened outcomes (Table 4). For example, young people aged 17-19 years using social media (eg, Facebook, Instagram, and Twitter) to connect with others were more likely to have anxiety but not depression at 3 years' follow-up [27]. A similar but older study found a worsening of depression outcomes with social networking interactions (ie, Facebook, Myspace, and texting) [26]. This relationship between negative interactions on social media and worsened outcomes was evident in both men and women, particularly in those also receiving low face-to-face social support [20]. Additional indicators of D-QSC that did not explicitly indicate effect on depression or anxiety were feeling validated [47], rapport [45], empathy [27,34], and emotional connection [48,50]. The indicators also improved well-being outcomes (Multimedia Appendix 4 [19,30-33,35,48,49,53]).

Figure 2. Forest plot showing the effect of social connection within digital interventions on depression outcomes. ES: effect size.

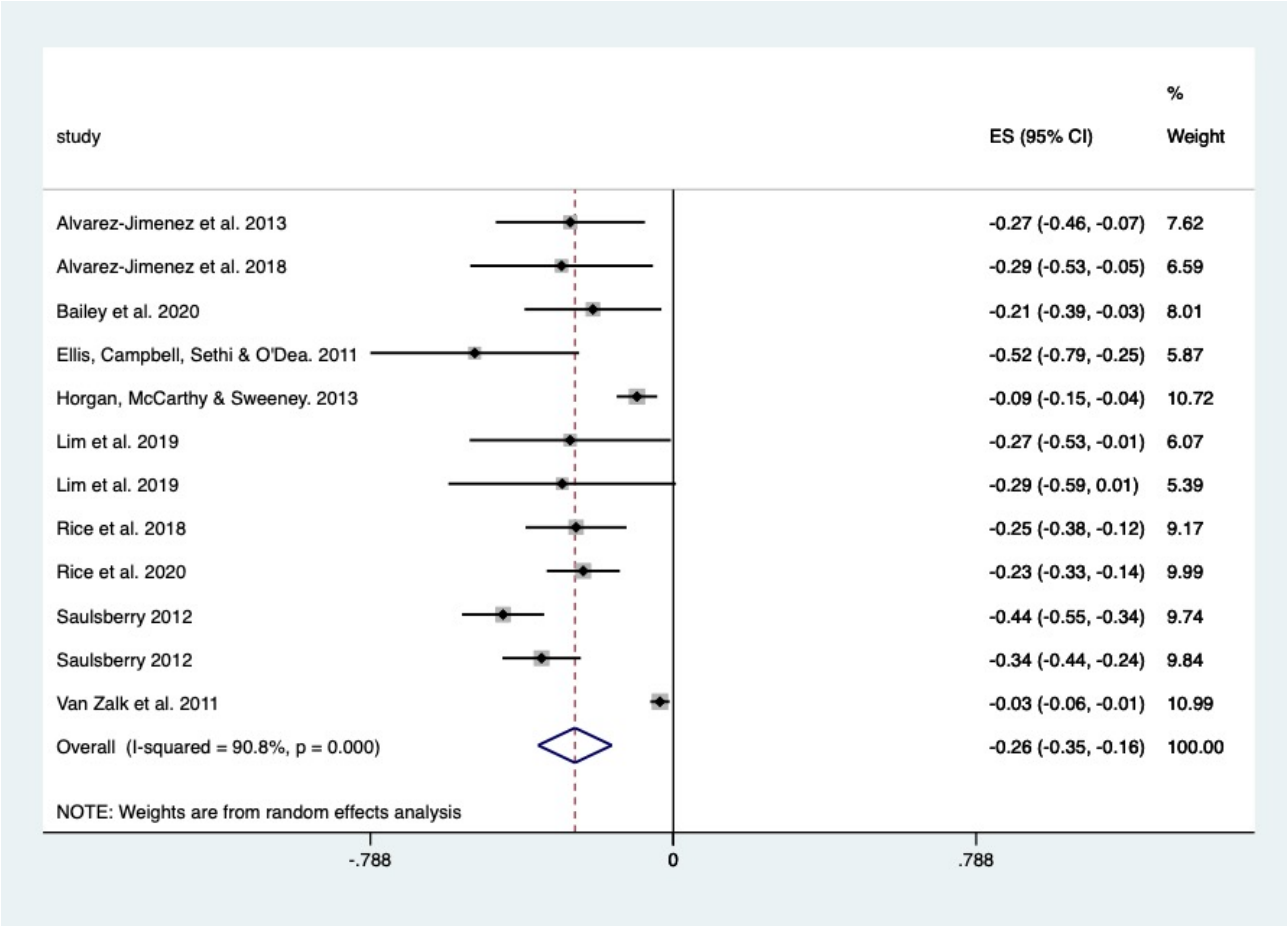
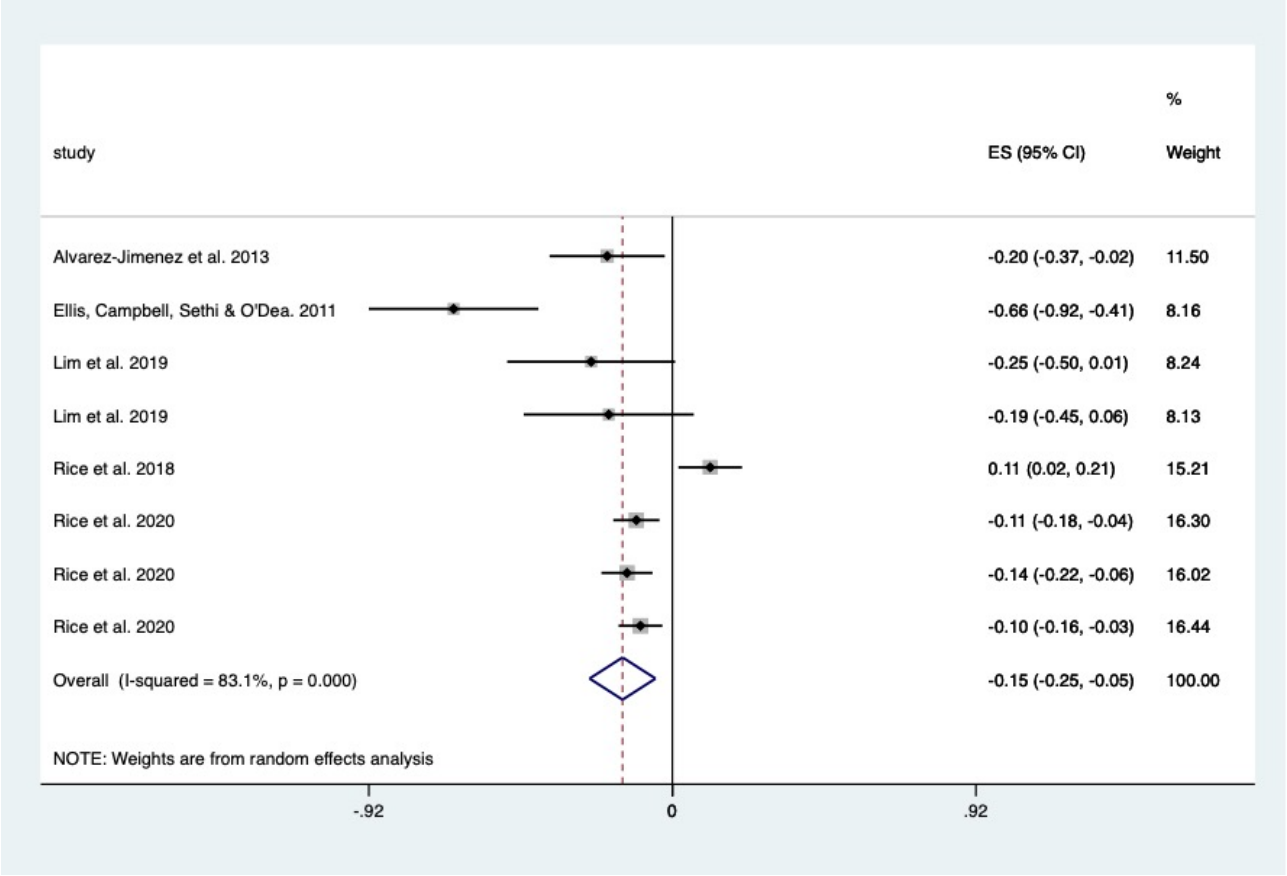


Figure 3. Forest plot showing the effect of social connection within digital interventions on anxiety outcomes. ES: effect size.



Digital Intervention Mechanisms That Facilitate QSC

Digital intervention mechanisms mainly helped facilitate QSC [18,20,25,30,35,43,44,46,50,51,55,56,58,59]. Forum moderation, confidentiality, ease of access, and anonymity supported by digital interventions were cited as valuable [18,43,50,56,58,59], facilitated open sharing in digital environments, and could lower inhibitions compared with face-to-face engagement [50,51]. The usual signals received during face-to-face interactions, such as body language or facial expression, were lost during digital interactions [52] and could impair the quality of interactions [26].

D-QSC was deemed more valuable when digital interactions were blended with face-to-face interactions [20,35,43]. For example, digital interactions were convenient and accessible, whereas face-to-face meetings helped maintain the connection. Higher levels of web-based social support were associated with increased symptoms of depression, specifically in both men and women and those who had little in-person social support [20]. One study found that participants were disconnected from the *real world* through high levels of web-based engagement [43]. Indeed, disconnection can have an impact on the interaction between peers and family and result in increased loneliness [19,32]. Other studies indicated that participants felt ignored [51], misunderstood [51], and had hurt feelings [43,49].

Participants also valued opportunities to support others [41,55], to connect with peers, and compare similar mental health experiences [25,54]. Some participants considered the networking component as the most helpful aspect of a moderated web-based social therapy tool, more helpful than the therapy itself [46]. Harassment was also identified as occurring frequently on the web. For example, a study reported this frequently among transgender adolescents [32].

Individual and Contextual Factors Influencing Mechanisms

Demographic and Personality Factors

The effect of D-QSC on depression outcomes differed across genders and personality variables. Social support from active Facebook use predicted a reduction in depression symptoms in girls but not in boys [37]. In another study, increased social support on the web and decreased offline social support was associated with increased depression symptoms across both genders [20]. Demographic (eg, personality type) and dynamic (eg, vulnerability level) characteristics were also reported to modulate the influence of D-QSC on depression and anxiety outcomes. Personality differences were only discussed in 5% (2/42) of the studies [41,48]. Chatting exclusively on the web predicted significantly improved depression [41] or psychological well-being [48] outcomes only in participants with more introverted personality traits.

Anxiety Versus Depression

D-QSC was more important for depression than for anxiety outcomes. For example, both web-based self-help cognitive behavioral therapy and peer support effectively reduced anxiety, but peer support was more effective in improving outcomes of depression [56]. Moreover, those with higher social anxiety had lower depression symptoms after corumination with a web-based

best friend [21]. In contrast, symptoms of depression predicted negative social networking interactions, which in turn resulted in higher symptoms of depression and anxiety [26]. An app designed to strengthen relationships and increase social connections for individuals with social anxiety disorder also improved symptoms of depression [57]. This effect lasted longer in participants without existing social anxiety disorder.

Offline–Web-Based Engagement

A cross-sectional study reported improved mood only for participants chatting with friends on the web who were also known offline; they were not web-exclusive friends [32]. Social web-based gamers who had lower depression and social anxiety on the web had higher QSC with friends offline [28].

Adapted Conceptual Framework

Stage 1

A total of 55 indicators were found from professionals (19/55, 35%), young people (19/55, 35%), and the literature (17/55, 30%; Multimedia Appendix 5). Social connectedness and social support were excluded because they directly described social connection and were not *attributes* of D-QSC (indicators). Of the 55 indicators, 5 (9%) were direct repeats and 5 (9%) were deemed not attributes of D-QSC.

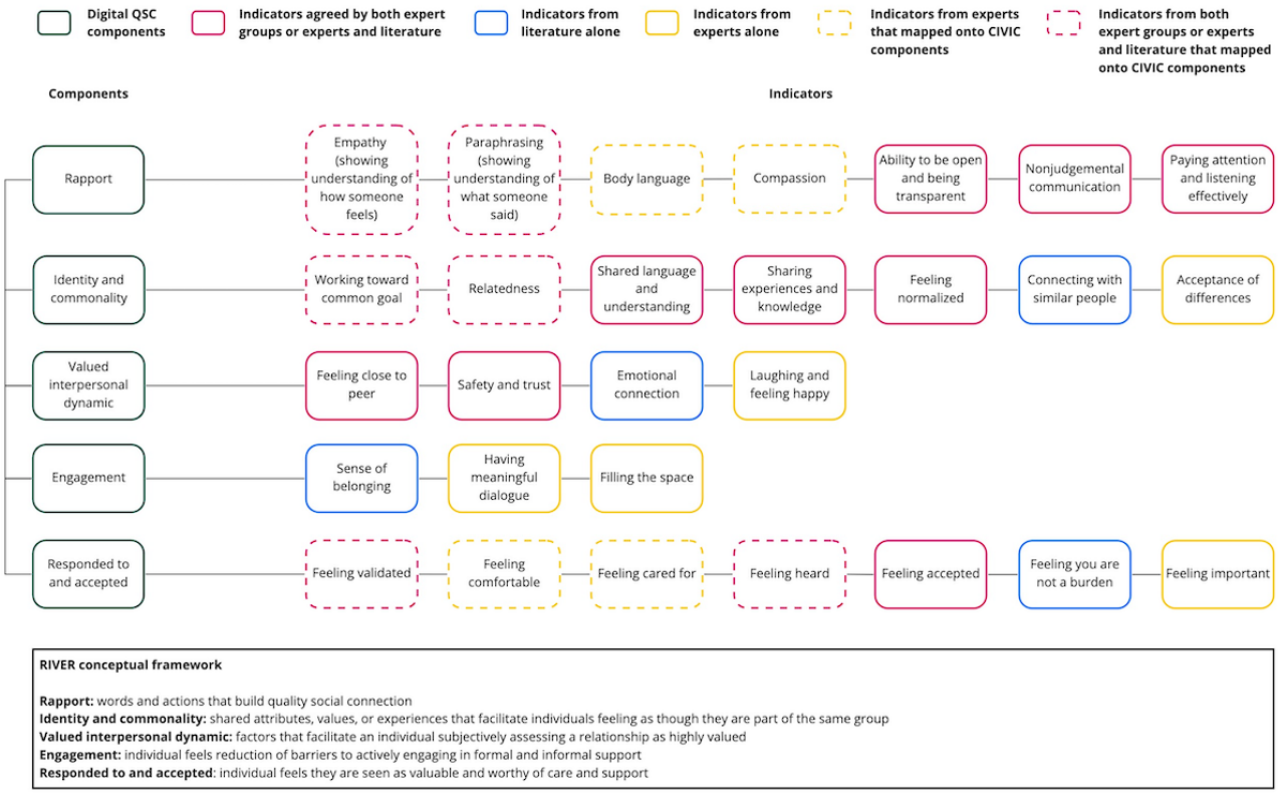
Stage 2

The remaining 45 indicators were grouped if conceptually similar or stood alone. For example, *Trust established* and *Trust*, as well as *Nonjudgmental* and *Not feeling judged*, were grouped, respectively. After the grouping, 30 indicators remained.

Stage 3

Of the 30 D-QSC indicators, 10 (33%) were initially mapped directly onto the preexisting CIVIC framework components (Figure 4). The remaining 67% (20/30) of indicators that did not map on directly were either merged with indicators that naturally went together, such as *Safety* and *trust*, or remained standalone indicators. Merged indicators (eg, *Safety and trust*, *Feeling close to peer*, and *Laughing and feeling happy*) and standalone indicators (eg, *Emotional connection*) were then loosely grouped and given new provisional component names (eg, Valued interpersonal dynamic) that suited the indicators' collective meaning.

Figure 4. Adapted RIVER (rapport, identity and commonality, valued interpersonal dynamic, engagement, and responded to and accepted) conceptual framework for quality social connection within digital interventions. CIVIC: Closeness, Identity and common bond, Valued relationships, Involvement, and Cared for and accepted; QSC: quality social connection.



Stage 4

Young experts reviewed the preliminary framework and identified 2 extra indicators that were deemed important (ie, *Feeling important* and *Acceptance of differences*) and added to the framework. Experts also helped to refine the wording or further merge indicators and components. For example, the *Identity and common bond* component was changed to *Identity and commonality*. There were 28 indicators across 5 components: rapport, identity and commonality, valued interpersonal dynamic, engagement, and responded to and accepted, given the acronym RIVER (Figure 4).

Discussion

Principal Findings

To our knowledge, this is the first systematic scoping review with meta-analysis to examine D-QSC as an *active ingredient* for depression and anxiety outcomes in young people. Usually conflated with quantity of social connections, QSC has now been comprehensively examined for its relevance to the mental health outcomes of digital interventions. We coproduced a conceptual framework of D-QSC for young people experiencing depression or anxiety that summarizes current understanding of component attributes or indicators. The RIVER framework comprises indicators relevant for establishing and assessing the presence of D-QSC. This can be characterized across 5 components. These components are interconnected and may not be exhaustive, but they provide a foundation for further work in this field to establish appropriate metrics for D-QSC.

D-QSC seems to help improve depression outcomes across most digital interventions. However, there is weaker evidence that D-QSC improves anxiety and well-being. There was also limited evidence of gaming, which was surprising considering that the participative nature with other users is at its core. D-QSC also worsened depression and anxiety outcomes in some instances, but this was often a result of negative interactions through social networking sites, which could be construed as a poor D-QSC. Few studies examined individual factors, contextual factors, or digital mechanisms that may modulate the impact of digital QSC on mental health outcomes. However, in the few studies that did report on mechanisms, a face-to-face connection before web-based support was an important consideration for improving outcomes. Furthermore, the impact of web-based support was modulated by the strength of offline connections.

Comparison With Prior Work

Reviews assessing the efficacy of digital mental health interventions for young people have found digital interventions to be as efficacious as, or sometimes more efficacious than, similar interventions delivered in person [65]. The strongest review to date that most closely relates to QSC collated measures of *social connectedness* to produce a conceptual framework of social connectedness in mental disorders (CIVIC) [17]. However, the review positioned social connectedness as the solution to loneliness and not as an *active ingredient* for the prevention and treatment of mental disorders (ie, depression and anxiety). Our work extended this framework to ensure that QSC was considered for digital interventions, for young people, and across different contexts. This process substantially expanded the elements indicating the development or presence of D-QSC and required redefining the CIVIC framework components to form the *RIVER framework of D-QSC indicators for young people*. Interestingly, the component of the CIVIC framework found to be most frequently assessed in current QSC metrics was *Identity and common bond*; this was the only component of our adapted framework not selected as a top priority in the context of digital interventions for anxiety and depression during review of the framework by 9 young people. This highlighted the need to develop improved metrics that are uniformly applied.

Strengths and Limitations

Young people with lived experience were involved at all review stages, including screening and interpretation. Dual independent review of the literature with people with direct experience of the review area helped us to overcome some limitations inherent in the current literature to gain better understanding of QSC and ensure accuracy of the screening. The insight from the young people during the data synthesis and interpretation stages helped to retain the data integrity. We have subsequently added to the limited evidence base for the impact of patient and public involvement throughout all stages of systematic and scoping reviews. Our adapted RIVER framework provides the foundation for future work to develop measures that would enable a developer, evaluator, or practitioner using digital interventions for mental health to assess the presence and degree of the QSC established.

The main limitation was that the studies did not control for a previous established connection offline before the D-QSC. Of the 42 studies, only 12 (29%) could be included in the meta-analysis because of a lack of measured effect sizes in previous work and heterogeneity across approaches, suggesting that the results should be interpreted with caution. Because of the scoping nature of the review, there was also statistical and methodological variability in the meta-analysis. Only manuscripts that were in English were included, which enables cultural bias. However, this was mitigated to an extent by working with young people and professionals from a variety of countries and cultural contexts to interpret the findings.

Clinical and Research Implications

QSC should be considered in the development and application of most digital interventions, particularly for depression. However, more research is needed to examine its impact within gaming platforms. In general, digital interventions mostly helped facilitate QSC; therefore, developers should consider factors such as user preference, anonymity, delivery medium, and content moderation. Initially, they should consider whether D-QSC is appropriate, depending on the target audience, and whether it will be important for engagement, or efficacy, or both. Further research is required to establish which individuals, conditions, and therapeutic mechanisms respond most strongly to D-QSC and what format is most appropriate. Clinical trials of any new digital intervention for mental health should control for previous face-to-face connections.

Future research should build upon our RIVER framework to further examine relationships among individual indicators of QSC, variations across different digital interventions, and the impact on outcomes across different user groups, particularly those in low- and middle-income countries. Factors that may mediate any causative relationships between QSC and mental health outcomes also deserve further attention. This work will inform the creation of standardized measures for D-QSC to evaluate its presence across different social settings. New measures should be developed to assess (1) attributes of a digital intervention that help or hinder good QSC and (2) the perceived value of a particular QSC for an individual and its relationship to outcomes within digital interventions. This work has value for development, regulation, and evaluation of digital mental

health interventions, as well as delineating helpful and harmful web-based interactions for young people, including social media. It will be vital to expand digital mental health care provision during the COVID-19 pandemic.

As the COVID-19 pandemic accelerates the shift to digital delivery of traditional therapy [13], clinicians should be trained in how to incorporate techniques for developing or maintaining D-QSC. Guidelines should be developed to ensure that moving face-to-face therapies to web-based spaces does not affect the QSC in the practitioner-patient dyad, and they should include strategies to improve connection on the web. Further clinical recommendations include a prioritization of video communication for web-based therapy to allow body language to be observed. However, anonymity can be beneficial to some users when first divulging sensitive mental health information. Blended care should enable patients to first meet their therapist in person, if desired, to facilitate QSC that can be translated to digital follow-ups.

Conclusions

In conclusion, D-QSC is important and an underconsidered component supporting engagement and efficacy for young people with depression and anxiety. In the wake of the COVID-19 pandemic, our work holds relevance as mental health needs rise and support will increasingly be provided on the web.

Acknowledgments

This work was funded by a Wellcome Trust Mental Health Priority Area *Active Ingredients* commission awarded to LD at Imperial College London.

The authors would like to thank Clara Nolan, Daniella Mallory, Emily Morse, Nathan Jackson, Nusaybah Choudhury, Ryan Vaughan, Tom Sharpe, Yasamin Neisi, and the other members of our UK Young Persons Advisory Group for all their excellent guidance and feedback throughout the project. The authors would also like to thank the Lancet Commission for Global Mental Health Young Leaders for their useful feedback from different global perspectives. Additional thanks go to the professional experts: Nicky Forsythe, Oliver Harrison, Heleen Riper, Victoria Hornby, Ariele Nobel, Anabelle Zhuang, Theresa Nguyen, Aliza Werner-Seidler, Tom Insel, Chris Martin, Jonny Benjamin, Peter Fonagy, and Peter Fuggle. Special thanks also go to Anna Lawrence-Jones and Catherine Kilkenny for their advice and guidance regarding patient and public involvement.

Authors' Contributions

LD and EL are joint first authors, led on conceptualization, and have contributed equally to writing this paper. LD led on patient and public involvement, synthesis, project management, and finalizing the paper. EL led on professional involvement. LR led on the scoping review, data extraction, and synthesis. EBH was involved in data screening, extraction, and quality assessment stages of the review. HA conducted the meta-analysis and advised on the review process. PA and GF provided project guidance. All authors reviewed and signed off the final paper.

Conflicts of Interest

HA is Chief Scientific Officer, Preemptive Medicine and Health Security at Flagship Pioneering. PA is a nonexecutive of West London NHS Trust, a mental health and community care provider.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 19 KB - jmir_v23i12e26584_app1.docx]

Multimedia Appendix 2
Stepwise approach to developing a full and objective search strategy.
[DOCX File , 24 KB - jmir_v23i12e26584_app2.docx]

Multimedia Appendix 3
Expanded data extraction and data quality assessment.
[PDF File (Adobe PDF File), 190 KB - jmir_v23i12e26584_app3.pdf]

Multimedia Appendix 4
Supplementary analyses.
[PDF File (Adobe PDF File), 109 KB - jmir_v23i12e26584_app4.pdf]

Multimedia Appendix 5
Indicators of social connection within digital interventions that contributed to the conceptual framework.
[PDF File (Adobe PDF File), 43 KB - jmir_v23i12e26584_app5.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
2. 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]
3. 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]
4. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. Lancet Psychiatry 2020 Oct;7(10):883-892 [FREE Full text] [doi: 10.1016/S2215-0366(20)30308-4] [Medline: 32707037]
5. Choi KW, Stein MB, Nishimi KM, Ge T, Coleman JR, Chen C, 23andMe Research Team, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, et al. An exposure-wide and mendelian randomization approach to identifying modifiable factors for the prevention of depression. Am J Psychiatry 2020 Oct 01;177(10):944-954. [doi: 10.1176/appi.ajp.2020.19111158] [Medline: 32791893]
6. Cruwys T, Alexander Haslam S, Dingle GA, Jetten J, Hornsey MJ, Desdemona Chong E, et al. Feeling connected again: interventions that increase social identification reduce depression symptoms in community and clinical settings. J Affect Disord 2014 Apr;159:139-146. [doi: 10.1016/j.jad.2014.02.019] [Medline: 24679402]
7. Therapeutic alliance—APA dictionary of psychology. American Psychology Association. 2020. URL: https://dictionary.apa.org/therapeutic-alliance [accessed 2021-11-19]
8. Murphy R, Hutton P. Practitioner review: therapist variability, patient-reported therapeutic alliance, and clinical outcomes in adolescents undergoing mental health treatment - a systematic review and meta-analysis. J Child Psychol Psychiatry 2018 Jan 06;59(1):5-19. [doi: 10.1111/jcpp.12767] [Medline: 28681928]
9. Pihlaja S, Stenberg J, Joutsenniemi K, Mehik H, Ritola V, Joffe G. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders - a systematic review. Internet Interv 2018 Mar;11:1-10 [FREE Full text] [doi: 10.1016/j.invent.2017.11.005] [Medline: 30135754]
10. Kellezi B, Wakefield JR, Stevenson C, McNamara N, Mair E, Bowe M, et al. The social cure of social prescribing: a mixed-methods study on the benefits of social connectedness on quality and effectiveness of care provision. BMJ Open 2019 Nov 14;9(11):e033137 [FREE Full text] [doi: 10.1136/bmjopen-2019-033137] [Medline: 31727668]
11. Levula A, Wilson A, Harré M. The association between social network factors and mental health at different life stages. Qual Life Res 2016 Jul;25(7):1725-1733. [doi: 10.1007/s11136-015-1200-7] [Medline: 26669317]
12. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010 Jul 27;7(7):e1000316 [FREE Full text] [doi: 10.1371/journal.pmed.1000316] [Medline: 20668659]
13. Chang BP, Kessler RC, Pincus HA, Nock MK. Digital approaches for mental health in the age of covid-19. BMJ 2020 Jun 29;369:m2541. [doi: 10.1136/bmj.m2541] [Medline: 32601049]
14. 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]
15. Sanger E. Social networking in mental health interventions for adolescents. Perspect Public Health 2020 Oct 01:1757913920924431 [FREE Full text] [doi: 10.1177/1757913920924431] [Medline: 33000675]
16. Garett R, Lord LR, Young SD. Associations between social media and cyberbullying: a review of the literature. Mhealth 2016 Dec 19;2:46 [FREE Full text] [doi: 10.21037/mhealth.2016.12.01] [Medline: 28293616]
17. Hare-Duke L, Dening T, de Oliveira D, Milner K, Slade M. Conceptual framework for social connectedness in mental disorders: systematic review and narrative synthesis. J Affect Disord 2019 Feb 15;245:188-199. [doi: 10.1016/j.jad.2018.10.359] [Medline: 30396057]
18. Horgan A, Sweeney J. Young students' use of the internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
19. Sharabi A, Margalit M. Virtual friendships and social distress among adolescents with and without learning disabilities: the subtyping approach. Eur J Spec Needs Educ 2011 Aug;26(3):379-394. [doi: 10.1080/08856257.2011.595173]
20. Yeh YC, Ko HC, Wu JY, Cheng CP. Gender differences in relationships of actual and virtual social support to internet addiction mediated through depressive symptoms among college students in Taiwan. Cyberpsychol Behav 2008 Aug;11(4):485-487. [doi: 10.1089/cpb.2007.0134] [Medline: 18721099]
21. Van Zalk N, Tillfors M. Co-rumination buffers the link between social anxiety and depressive symptoms in early adolescence. Child Adolesc Psychiatry Ment Health 2017;11:41 [FREE Full text] [doi: 10.1186/s13034-017-0179-y] [Medline: 28852420]
22. Bailey E, Alvarez-Jimenez M, Robinson J, D'Alfonso S, Nedeljkovic M, Davey CG, et al. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. Int J Environ Res Public Health 2020 Apr 03;17(7):2435 [FREE Full text] [doi: 10.3390/ijerph17072435] [Medline: 32260111]
23. McCloskey W, Iwanicki S, Lauterbach D, Giammittorio DM, Maxwell K. Are Facebook "friends" helpful? Development of a Facebook-based measure of social support and examination of relationships among depression, quality of life, and social support. Cyberpsychol Behav Soc Netw 2015 Sep;18(9):499-505. [doi: 10.1089/cyber.2014.0538] [Medline: 26348809]

24. Blackwell L, Romero S, Romero C, McLean K, Dawkins K, Hoag J. CFfone: a social networking site for adolescents and young adults with cf. In: Proceedings of the Pediatric Pulmonology - The 26th Annual North American Cystic Fibrosis Conference. 2012 Presented at: Pediatric Pulmonology - The 26th Annual North American Cystic Fibrosis Conference; Oct 11-13th, 2012; Orlando, Florida, USA p. A.
25. Canady VA. Survey explores social media, mental well-being among youth. Ment Health Wkly 2018 Jul 30;28(29):3-4. [doi: 10.1002/mhw.31535]
26. Feinstein BA, Bhatia V, Hershenberg R, Davila J. Another venue for problematic interpersonal behavior: the effects of depressive and anxious symptoms on social networking experiences. J Soc Clin Psychol 2012 Apr;31(4):356-382. [doi: 10.1521/jscp.2012.31.4.356]
27. Stockdale LA, Coyne SM. Bored and online: reasons for using social media, problematic social networking site use, and behavioral outcomes across the transition from adolescence to emerging adulthood. J Adolesc 2020 Feb;79:173-183. [doi: 10.1016/j.adolescence.2020.01.010] [Medline: 31978836]
28. Colder Carras M, Van Rooij AJ, Van de Mheen D, Musci R, Xue QL, Mendelson T. Video gaming in a hyperconnected world: a cross-sectional study of heavy gaming, problematic gaming symptoms, and online socializing in adolescents. Comput Human Behav 2017 Mar;68:472-479 [FREE Full text] [doi: 10.1016/j.chb.2016.11.060] [Medline: 28260834]
29. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
30. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
31. Rice S, O'Bree B, Wilson M, McEnery C, Lim MH, Hamilton M, et al. Leveraging the social network for treatment of social anxiety: pilot study of a youth-specific digital intervention with a focus on engagement of young men. Internet Interv 2020 Apr;20:100323 [FREE Full text] [doi: 10.1016/j.invent.2020.100323] [Medline: 32435600]
32. Sharabi A, Margalit M. The mediating role of internet connection, virtual friends, and mood in predicting loneliness among students with and without learning disabilities in different educational environments. J Learn Disabil 2011;44(3):215-227. [doi: 10.1177/0022219409357080] [Medline: 20375289]
33. Poppelaars M, Lichtwarck-Aschoff A, Kleinjan M, Granic I. The impact of explicit mental health messages in video games on players' motivation and affect. Comput Hum Behav 2018 Jun;83:16-23. [doi: 10.1016/j.chb.2018.01.019]
34. Felnhofer A, Kafka JX, Hlavacs H, Beutl L, Kryspin-Exner I, Kothgassner OD. Meeting others virtually in a day-to-day setting: investigating social avoidance and prosocial behavior towards avatars and agents. Comput Hum Behav 2018 Mar;80:399-406. [doi: 10.1016/j.chb.2017.11.031]
35. Liu CY, Yu CP. Can Facebook use induce well-being? Cyberpsychol Behav Soc Netw 2013 Sep;16(9):674-678. [doi: 10.1089/cyber.2012.0301] [Medline: 24028138]
36. Cole DA, Nick EA, Zelkowitz RL, Roeder KM, Spinelli T. Online social support for young people: does it recapitulate in-person social support; can it help? Comput Human Behav 2017 Mar;68:456-464 [FREE Full text] [doi: 10.1016/j.chb.2016.11.058] [Medline: 28993715]
37. Frison E, Eggermont S. Exploring the relationships between different types of facebook use, perceived online social support, and adolescents' depressed mood. Soc Sci Comput Rev 2015 Jan 21;34(2):153-171. [doi: 10.1177/0894439314567449]
38. Mikami AY, Szwedo DE, Allen JP, Evans MA, Hare AL. Adolescent peer relationships and behavior problems predict young adults' communication on social networking websites. Dev Psychol 2010 Jan;46(1):46-56 [FREE Full text] [doi: 10.1037/a0017420] [Medline: 20053005]
39. Ozcan NK, Buzlu S. Internet use and its relation with the psychosocial situation for a sample of university students. Cyberpsychol Behav 2007 Dec;10(6):767-772. [doi: 10.1089/cpb.2007.9953] [Medline: 18085963]
40. Saulsberry A, Marko-Holguin M, Blomeke K, Hinkle C, Fogel J, Gladstone T, et al. Randomized clinical trial of a primary care internet-based intervention to prevent adolescent depression: one-year outcomes. J Can Acad Child Adolesc Psychiatry 2013 May;22(2):106-117 [FREE Full text] [Medline: 23667356]
41. Van Zalk MH, Branje SJ, Denissen J, Van Aken MA, Meeus WH. Who benefits from chatting, and why? The roles of extraversion and supportiveness in online chatting and emotional adjustment. Pers Soc Psychol Bull 2011 Sep;37(9):1202-1215. [doi: 10.1177/0146167211409053] [Medline: 21673194]
42. Wright KB, Rosenberg J, Egbert N, Ploeger NA, Bernard DR, King S. Communication competence, social support, and depression among college students: a model of facebook and face-to-face support network influence. J Health Commun 2013;18(1):41-57. [doi: 10.1080/10810730.2012.688250] [Medline: 23030518]
43. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young people's response to six smartphone apps for anxiety and depression: focus group study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
44. Radovic A, DeMand AL, Gmelin T, Stein BD, Miller E. SOVA: design of a stakeholder informed social media website for depressed adolescents and their parents. J Technol Hum Serv 2018;35(3):169-182 [FREE Full text] [doi: 10.1080/15228835.2017.1347552] [Medline: 29743822]

45. Clarke CS. Telepsychiatry in Asperger's syndrome. Ir J Psychol Med 2018 Dec;35(4):325-328. [doi: 10.1017/ipm.2017.19] [Medline: 30501669]
46. Santesteban-Echarri O, Rice S, Wadley G, Lederman R, D'Alfonso S, Russon P, et al. A next-generation social media-based relapse prevention intervention for youth depression: qualitative data on user experience outcomes for social networking, safety, and clinical benefit. Internet Interv 2017 Sep 29;9:65-73 [FREE Full text] [doi: 10.1016/j.invent.2017.06.002] [Medline: 30135839]
47. Selkie E, Adkins V, Masters E, Bajpai A, Shumer D. Transgender adolescents' uses of social media for social support. J Adolesc Health 2020 Mar;66(3):275-280. [doi: 10.1016/j.jadohealth.2019.08.011] [Medline: 31690534]
48. Dolev-Cohen M, Barak A. Adolescents' use of instant messaging as a means of emotional relief. Comput Human Behav 2013 Jan;29(1):58-63. [doi: 10.1016/j.chb.2012.07.016]
49. Bhuvaneswar CG, Gutheil TG. E-mail and psychiatry: some psychotherapeutic and psychoanalytic perspectives. Am J Psychother 2008;62(3):241-261. [doi: 10.1176/appi.psychotherapy.2008.62.3.241] [Medline: 18846971]
50. van Rensburg SH, Klingensmith K, McLaughlin P, Qayyum Z, van Schalkwyk GI. Patient-provider communication over social media: perspectives of adolescents with psychiatric illness. Health Expect 2016 Feb;19(1):112-120 [FREE Full text] [doi: 10.1111/hex.12334] [Medline: 25581724]
51. Dhesi M. A qualitative study to investigate in what ways are the distinctive features of synchronous text- based counselling experienced as being helpful and/or unhelpful by young people? University of Roehampton. 2019. URL: https://pure.roehampton.ac.uk/portal/en/studentTheses/a-qualitative-study-to-investigate-in-what-ways-are-the-distincti [accessed 2021-11-21]
52. Siriaraya P, Tang C, Ang CS, Pfeil U, Zaphiris P. A comparison of empathic communication pattern for teenagers and older people in online support communities. Behav Inf Technol 2011 Sep;30(5):617-628. [doi: 10.1080/0144929x.2011.582146]
53. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. J Adolesc 2017 Feb;55:5-15 [FREE Full text] [doi: 10.1016/j.adolescence.2016.12.002] [Medline: 27997851]
54. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89. [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]
55. Campbell A, Ridout B, Amon K, Navarro P, Collyer B, Dalgleish J. A customized social network platform (Kids Helpline Circles) for delivering group counseling to young people experiencing family discord that impacts their well-being: exploratory study. J Med Internet Res 2019 Dec 20;21(12):e16176 [FREE Full text] [doi: 10.2196/16176] [Medline: 31859671]
56. Ellis L, Campbell A, Sethi S, O'Dea B. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. J Cyber Ther Rehabil 2011;4:461-467 [FREE Full text]
57. Lim MH, Rodebaugh TL, Eres R, Long KM, Penn DL, Gleeson JF. A pilot digital intervention targeting loneliness in youth mental health. Front Psychiatry 2019 Aug 23;10:604 [FREE Full text] [doi: 10.3389/fpsyt.2019.00604] [Medline: 31507469]
58. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
59. Chyzzy B, Nelson LE, Stinson J, Vigod S, Dennis C. Adolescent mothers' perceptions of a mobile phone-based peer support intervention. Can J Nurs Res 2020 Jun;52(2):129-138. [doi: 10.1177/0844562120904591] [Medline: 32036679]
60. Hawker S, Payne S, Kerr C, Hardey M, Powell J. Appraising the evidence: reviewing disparate data systematically. Qual Health Res 2002 Nov;12(9):1284-1299. [doi: 10.1177/1049732302238251] [Medline: 12448672]
61. 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]
62. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
63. Oxford English Dictionary Home. Oxford English Dictionary. URL: https://www.oed.com [accessed 2021-12-07]
64. Budgeting for involvement: Practical advice on budgeting for actively involving the public in research studies. National Institute for Health Research. 2013. URL: https://www.invo.org.uk/wp-content/uploads/2013/07/INVOLVEMHRNBudgeting09Jul2013.pdf [accessed 2021-12-07]
65. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]

Abbreviations

CIVIC: closeness, identity and common bond, valued relationships, involvement, and cared for and accepted
D-QSC: quality social connection within digital interventions
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QSC: quality social connection

RIVER: rapport, identity and commonality, valued interpersonal dynamic, engagement, and responded to and accepted
YPAG: Young Persons Advisory Group

Edited by R Kukafka, G Eysenbach; submitted 18.12.20; peer-reviewed by P Copanitsanou, S Rennick-Egglestone, A Jolliff; comments to author 03.02.21; revised version received 30.03.21; accepted 14.10.21; published 17.12.21.

Please cite as:
Dewa LH, Lawrance E, Roberts L, Brooks-Hall E, Ashrafian H, Fontana G, Aylin P
Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis
J Med Internet Res 2021;23(12):e26584
URL: https://www.jmir.org/2021/12/e26584
doi:10.2196/26584
PMID:34927592

Review

The Long-Term Effectiveness of Internet-Based Interventions on Multiple Health Risk Behaviors: Systematic Review and Robust Variance Estimation Meta-analysis

Flora Tzelepis[1,2,3], PhD; Aimee Mitchell[1,2], BPsych (Hons); Louise Wilson[1,2], BNutrDiet (Hons); Emma Byrnes[1,3], BSocSc (Hons); Alexandra Haschek[4,5], BPsych (Hons); Lucy Leigh[3], PhD; Christopher Oldmeadow[3], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Hunter New England Local Health District, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

[4]School of Psychology and Public Health, La Trobe University, Bundoora, Australia

[5]The Cooperative Research Centre for Living with Autism, Brisbane, Australia

Corresponding Author:
Flora Tzelepis, PhD
School of Medicine and Public Health
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 2 49246275
Fax: 61 2 4924 6490
Email: flora.tzelepis@newcastle.edu.au

Abstract

Background: Smoking tobacco, poor nutrition, risky alcohol use, and physical inactivity (SNAP) behaviors tend to cluster together. Health benefits may be maximized if interventions targeted multiple health risk behaviors together rather than addressing single behaviors. The internet has wide reach and is a sustainable mode for delivery of interventions for multiple health behaviors. However, no systematic reviews have examined the long-term effectiveness of internet-based interventions on any combination of or all SNAP behaviors in adults aged 18 years or older.

Objective: This systematic review examined, among adults (aged ≥18 years), the effectiveness of internet-based interventions on SNAP behaviors collectively in the long term compared with a control condition.

Methods: The electronic databases Medline, PsycINFO, Embase, CINAHL, and Scopus were searched to retrieve studies describing the effectiveness of internet-based interventions on ≥2 SNAP behaviors published by November 18, 2019. The reference lists of retrieved articles were also checked to identify eligible publications. The inclusion criteria were randomized controlled trials or cluster randomized controlled trials with adults examining an internet-based intervention measuring the effect on ≥2 SNAP behaviors at least 6 months postrecruitment and published in English in a peer-reviewed journal. Two reviewers independently extracted data from included studies and assessed methodological quality using the Quality Assessment Tool for Quantitative Studies. A robust variance estimation meta-analysis was performed to examine the long-term effectiveness of internet-based interventions on all 4 SNAP risk behavior outcomes. All SNAP outcomes were coded so they were in the same direction, with higher scores equating to worse health risk behaviors.

Results: The inclusion criteria were met by 11 studies: 7 studies measured the effect of an internet-based intervention on nutrition and physical activity; 1 study measured the effect on smoking, nutrition, and physical activity; and 3 studies measured the effect on all SNAP behaviors. Compared with the control group, internet-based interventions achieved an overall significant improvement across all SNAP behaviors in the long term (standardized mean difference –0.12 [improvement as higher scores = worse health risk outcomes], 95% CI –0.19 to –0.05; I^2=1.5%, *P*=.01). The global methodological quality rating was “moderate” for 1 study, while the remaining 10 studies were rated as “weak.”

Conclusions: Internet-based interventions were found to produce an overall significant improvement across all SNAP behaviors collectively in the long term. Internet-based interventions targeting multiple SNAP behaviors have the potential to maximize long-term improvements to preventive health outcomes.

(J Med Internet Res 2021;23(12):e23513) doi:10.2196/23513

KEYWORDS

internet; multiple health behaviors; tobacco; nutrition; alcohol; physical activity

Introduction

Smoking tobacco, poor nutrition, risky alcohol use, and physical inactivity (SNAP) are modifiable risk factors for chronic diseases such as heart disease, stroke, cancer, and diabetes [1]. Individuals who engage in all 4 SNAP behaviors, compared with 0, have an increased risk of mortality, equivalent to 14 years of aging [2]. Evidence has shown that SNAP behaviors tend to cluster together [3-5], suggesting a holistic approach for interventions to modify multiple health behaviors collectively rather than single behaviors individually may be beneficial. Multiple health behavior interventions target 2 or more health behaviors either sequentially or simultaneously [6]. Advantages of improving multiple health behaviors include maximizing health benefits [7], greater reduction in medical costs [8], and successfully modifying one behavior may increase confidence or motivation to change other health behaviors [7,9].

The internet is accessible globally and is a sustainable mode for the delivery of interventions for multiple health behaviors [10]. There are more than 4 billion internet users worldwide [10]; therefore, internet-based interventions have the potential to reach large numbers of people. Other advantages of internet-based interventions include that users can access information any time [11,12] as well as a low-cost modality for information delivery [11,12] and allowing for privacy, confidentiality [11], and long-term use [12,13]. Internet-based interventions may be interactive [11-13] and incorporate behavior change techniques such as individually tailored information [11,13], goal setting [12-14], self-monitoring [12-14], personalized and normative feedback [11,12,14], and progress tracking [12,13]. Internet-based interventions may also reduce health inequalities by improving access to services, for example among individuals who live in rural and remote areas or have significant mobility issues [13,15].

Existing systematic reviews have examined the effectiveness of behavioral interventions on multiple health risk behaviors [16,17]. A systematic review of nonpharmacologic interventions on multiple health risk behaviors found modest improvements in fruit and vegetable intake, physical activity, reduced fat intake, and reduced smoking [16]. Furthermore, another systematic review examining the efficacy of apps in children, adolescents, and adults reported that 41% of multiple health behavior interventions showed significant between-group improvements in behaviors [17]. However, only 2 systematic reviews have specifically examined the effectiveness of internet-based interventions on 2 or more SNAP behaviors in adult populations [18,19]. The review by Norman et al [18] focused on interventions for nutrition and physical activity but not tobacco use and alcohol intake. Of the 17 studies targeting multiple behaviors, 6 studies favored an internet-based intervention for increasing physical activity, and 6 studies favored an internet-based intervention for changing nutrition behaviors [18]. However, this systematic review was not restricted to adult populations, and findings from children and adolescents were included in the synthesis of findings [18]. Furthermore, short-term follow-up assessments were contained within this systematic review, and many studies did not report the effect of the internet-based intervention on nutrition and physical activity in the long term [18]. In the second systematic review, Oosterveen and colleagues [19] examined the effectiveness of internet-based interventions on combinations of all SNAP behaviors but included young adults aged 18 years to 35 years only. This systematic review identified only 2 studies with young adults targeting nutrition and physical activity behaviors that included a long-term follow-up (ie, 6 months or longer) [19]. To our knowledge, there is no systematic review that has examined the long-term effectiveness of internet-based interventions on any combination of or all SNAP behaviors in adults aged 18 years or older. Further critical review of the evidence is therefore needed to understand whether internet-based interventions are effective in improving multiple SNAP behaviors in the long term.

This systematic review aimed to examine the effectiveness of internet-based interventions on multiple SNAP health risk behaviors in the long term compared with a control condition.

Methods

Search Strategy and Selection Criteria

The electronic databases Medline, PsycINFO, Embase, CINAHL, and Scopus were searched to retrieve studies describing the effectiveness of internet-based interventions on 2 or more SNAP behaviors published by November 18, 2019. The following combinations of keywords were used: (multiple health behavio* or multiple behavio* or multiple risk* or multiple health* or smok* or tobacco or alcohol or diet* or nutrition or exercise or physical activity or fruit* or vegetable*) AND (internet or web* or online or on-line) AND (trial* or RCT* or random*). The reference lists of retrieved articles were also checked to identify any additional eligible publications.

The inclusion criteria were studies (1) that reported randomized controlled trials (RCTs) or cluster RCTs of internet-based interventions for ≥2 SNAP behaviors as either the sole intervention or an adjunct to written materials, (2) with adults aged 18 years or older, (3) that reported outcomes for ≥2 SNAP behaviors at least 6 months postrecruitment, (4) that had a no-intervention control group or the control group received information either in hard copy or information unrelated to

SNAP via a website, and (5) in the English language in a peer-reviewed journal.

Publications were excluded if (1) they did not report the outcomes of an RCT or cluster RCT (eg, systematic reviews, commentaries); (2) they examined only 1 SNAP health behavior; (3) they included special populations only such as people with chronic conditions (eg, cancer, diabetes) or pregnant women (this criterion was chosen because people with chronic conditions may differ in their motivation and capacity to change behaviors compared with those without chronic conditions and is consistent with the criterion set in another systematic review that examined multiple health risk behaviors [16]); (4) were conducted with people under 18 years of age; (5) outcome measures were not related to SNAP (eg, blood pressure); (6) the internet-based intervention was part of a multicomponent approach that included other modes of support (eg, face-to-face, telephone); (7) there was no control arm, and instead, comparisons were made with other interventions (eg, face-to-face support); and (8) SNAP outcomes were measured before 6 months postrecruitment.

Selection of Eligible Studies

This systematic review was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20]. All records identified in each electronic database were imported into Endnote, and duplicates were removed. Titles, abstracts, and full texts of each reference were independently screened in duplicate by 2 reviewers (LW and FT or AM or EB) to determine if eligibility criteria were met. Full-text articles were retrieved when eligibility could not be determined from the title and abstract screening.

Study and Sample Characteristics

Data were independently extracted from the included studies by 2 authors (AM and AH or LW). A third author (FT) resolved any inconsistencies in data extraction. The study and sample characteristics extracted from eligible publications included authors and year of publication, country, years that data were collected, setting, sample characteristics (eg, mean age, gender, education, employment status), recruitment method, eligibility criteria, treatment conditions (relevant arms only; ie, internet-based intervention and control arms), internet-based intervention received (eg, duration, number of modules), retention rate at follow-up, SNAP measures, SNAP outcomes at 6 months of follow-up or later, and costs.

The outcomes extracted for each health behavior were any measure of (1) tobacco smoking (eg, current tobacco smoking, point prevalence abstinence, or prolonged abstinence), (2) nutrition (eg, number of daily serves of fruit and vegetables, dietary score), (3) alcohol consumption (eg, number of alcoholic drinks per day), and (4) physical activity (eg, moderate to vigorous physical activity, metabolic equivalent of task [MET] minutes per week).

Methodological Quality Assessment

The Quality Assessment Tool for Quantitative Studies developed by the Effective Public Health Practice Project was used to assess methodological quality [21]. This tool was chosen because it has demonstrated content validity, construct validity, and test-retest reliability [22] and has been shown to have higher interrater reliability than the Cochrane Collaboration Risk of Bias Tool [23]. The Quality Assessment Tool for Quantitative Studies allows randomized trials to be rated on 6 components: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, and (6) withdrawals and dropouts. Each study was rated as "strong," "moderate," or "weak" for each component. An overall global rating was then assigned to each study, with studies classified as "strong" (no weak ratings), "moderate" (1 weak rating), or "weak" (2 or more weak ratings). The Quality Assessment Tool for Quantitative Studies Dictionary was used to make judgments about each of the 6 components [24].

The methodological quality of included studies was rated by 2 authors (FT and AM). Any discrepancies were discussed between these authors until consensus was reached. When rating the data collection methods, the measures for "all" SNAP outcomes needed to be shown to be valid and reliable for the data collection methods to be rated as strong. For instance, if smoking cessation was measured via biochemical validation and physical activity assessed via pedometers, the data collection methods were rated as strong. However, if smoking cessation was measured via a self-reported measure with no information about its psychometric properties and pedometers were used to assess physical activity, the data collection methods were rated as weak because "all" SNAP measures were not shown to be reliable and valid.

Robust Variance Estimation Meta-analysis

A robust variance estimation meta-analysis was performed using the R package *robumeta*. All SNAP behaviors were coded so they were in the same direction, with higher scores equating to worse health risk outcomes. Where a study measured the outcome at multiple time points (eg, 6 months and 12 months), data from the longer-term follow-up was included in the meta-analysis. The SNAP outcomes from each study were converted into Cohen d (standardized mean differences [SMDs]) and the corresponding variance [25]. Robust variance estimation meta-analysis was then performed on the SMDs (and variances), using the R package *robumeta*. A common within-study correlation (rho) of 0.8 was assumed, and sensitivity analyses were performed to determine the sensitivity of the results to this assumption (by looking at the results across various rho equal to 0, 0.2, 0.4, 0.6, or 1.0). Heterogeneity was measured using the I^2 statistic [26].

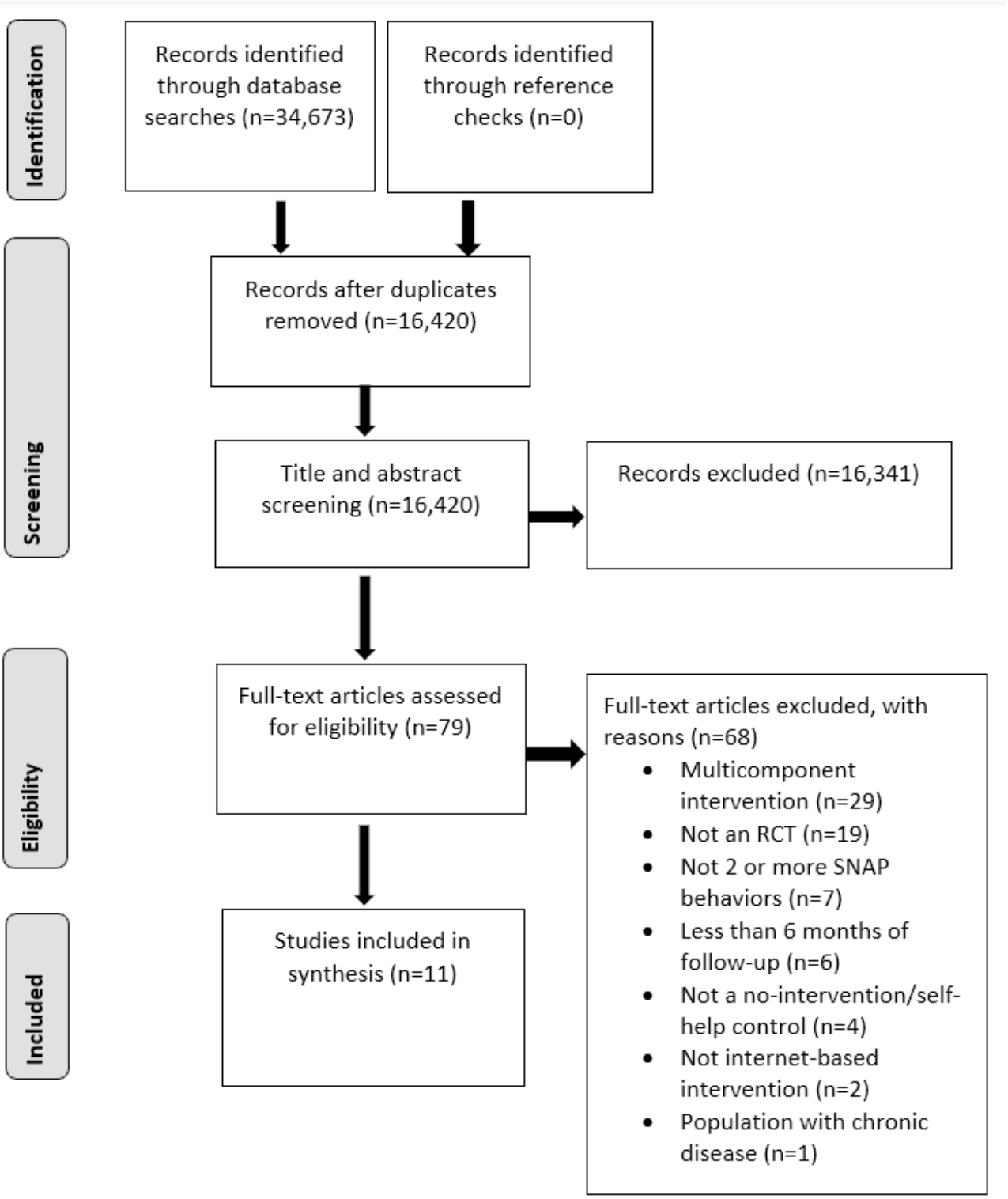
Results

Search Results

The PRISMA flow diagram outlining the study selection process is presented in Figure 1. A total of 34,673 records were identified from the electronic database searches. After the removal of duplicates, 16,420 records had their title and/or abstract screened, and 16,341 did not meet the inclusion criteria. The full texts of the remaining 79 articles were retrieved for further review, and 68 articles were excluded for the following

reasons: multicomponent intervention (n=29), not an RCT (n=19), did not measure 2 or more SNAP behaviors (n=7), less than 6 months of follow-up (n=6), did not include a no-intervention or self-help control group (n=4), not an internet-based intervention (n=2), and population with chronic disease (n=1). The inclusion criteria for this review were met by 11 studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of the screening and selection process. RCT: randomized controlled trial.



Study Characteristics of Internet-Based Interventions for Multiple Health Risk Behaviors

Studies That Examined Nutrition and Physical Activity

As shown in Multimedia Appendix 1, the effectiveness of an internet-based intervention on nutrition and physical activity was examined by 7 studies [27-33]; 5 studies were conducted in the United States [27,29,30,32,33], and 1 study each was conducted in the United Kingdom [31] and Australia [28]. Recruitment occurred from the community in 3 studies [28,32,33], from health care settings in 2 studies [27,31], and from universities in 2 studies [29,30]. In 6 studies, participants were recruited via various advertisements (eg, website, emails, newspaper, flyers, posters, radio) [27,28,30-33], while 2 studies used sign-up tables at universities [29,30]. The sample size ranged from 121 [27] to 1071 [33]. Only men were included in 2 studies [28,32], and only women were included in 1 study [27]. The percentage of women ranged from 56.3% (268/476) to 77% (171/221) in the remaining studies [29-31,33].

The interventions included web-based learning activities, modules, or tutorials [27,30,32,33]; self-monitoring [28,32]; educational materials [28,29,31]; social support [28]; tailored information [27,30-33]; feedback [32]; and goal setting [32,34]. The duration of engagement with the internet-based interventions ranged from 2 sessions [29] to a 12-month trial period [31]. Nutrition and physical activity were measured via self-reported measures in 6 studies [27-32], while 1 study used self-reported measures for nutrition and pedometers for physical activity [33]. Nutrition and physical activity were assessed at 6 months in 4 studies (range of retention of 291/441, 66.0% to 105/121, 86.8% [27,29,31,32]), 7 months in 1 study (retention: 950/1071, 88.7% [33]), 9 months in 1 study (retention: 148/317, 46.7% [28]), 12 months in 2 studies (retention: 131/221, 59.3% and 309/441, 70.1% [31,32]), 15 months in 1 study (retention: 1126/1689, 66.7% [30]), and 16 months in 1 study (retention: 935/1071, 87.3% [33]).

Studies That Examined Smoking, Nutrition, and Physical Activity

Multimedia Appendix 2 describes the study characteristics of the US study examining the effectiveness of an internet-based intervention on smoking, nutrition, and physical activity [34]. This study recruited 423 university staff (female: 347/423, 82%; mean age 51 years) via announcements on staff listservs, targeted emails, recruitment tables at events, and flyers [34]. The RealAge internet program generated individual risk profiles and allowed users to select behaviors to change and create plans to meet behavioral goals [34]. Self-reported measures assessed smoking cessation, nutrition, and physical activity at 6-month (retention: 360/423, 85.1%) and 12-month (retention: 367/423, 86.8%) follow-ups [34].

Studies That Examined Smoking, Nutrition, Alcohol, and Physical Activity

As outlined in Multimedia Appendix 3, the effectiveness of internet-based interventions on smoking, nutrition, alcohol, and physical activity was assessed by 3 studies [35-37]. All studies were conducted in the United Kingdom [35-37]; 2 studies were undertaken in the university setting [35,36], and 1 study was conducted in the community [37]. In 2 studies, incoming undergraduate students were recruited via an email invitation [35,36], while online and print advertisements were used to recruit participants in the community study [37]. Across the 3 studies, the sample size ranged from 100 [37] to 2621 [35]. Most participants were women in all the studies (range: 1447/2614, 55.4% to 82/100, 82% [35-37]), and the mean age ranged from 18.8 years [35] to 39 years [37]. The U@Uni [36] and U@Uni:LifeGuide [35] internet-based interventions included a profile page containing self-affirmation manipulation, theory-based messages for each SNAP behavior, and a planner to form implementation intentions. The HealthyValues Healthy Eating program targeted motivation, volition, and maintenance and included weekly tasks for 24 weeks [37]. All studies measured SNAP outcomes via self-reported measures at a 6-month follow-up [35-37]. The retention rates at 6 months were 41.2% (1079/2621) [35], 63.2% (913/1445) [36], and 95% (95/100) [37], respectively.

Long-Term Effectiveness of Internet-Based Interventions Across SNAP Risk Behaviors

The robust variance estimation meta-analysis found that, compared with the control group, internet-based interventions achieved an overall significant improvement across all SNAP behaviors in the long-term (SMD –0.12 [improvement as higher scores = worse health risk outcomes], 95% CI –0.19 to –0.05; I^2=1.5%, *P*=.01) [27,29,30,32,34-37]. Heterogeneity was low (I^2=1.5%), and the tau-square (the extent of variation due to between-study variance) was low, at <0.001. The sensitivity analysis showed that the model results did not vary greatly across different values of within-study correlation (rho).

Methodological Quality Assessment

Table 1 presents the methodological quality ratings for the 6 components and the global rating as assessed via the Quality Assessment Tool for Quantitative Studies [21]. In terms of selection bias, all studies were rated "weak," as 9 of these studies recruited volunteers who responded to advertisements or sign-up tables [27-34,37], and 2 studies sent emails to all incoming undergraduate students but recruited less than half of those approached [35,36]. All included studies were RCTs or cluster RCTs and were rated as "strong" in terms of study design [27-37]. With regards to confounders, 8 studies were rated as "strong" because there were no between-group differences at baseline [31,32,34,35,37] or adjustments for baseline characteristics were made during analysis [27,29,36], while 1 study was rated as "moderate," as stratification attempted to balance baseline characteristics across groups [33], and 2 studies were rated as "weak," as no adjustments were made during analysis for baseline characteristics that differed between the groups [28,30]. For blinding, 9 studies were rated as "weak" because the assessors and participants were not blinded [28,31,34] or there was no information about blinding [27,29,30,33,35,36], while 2 studies were rated as "moderate" because only the assessors were blinded to the condition [32,37]. For data collection methods, 8 studies were rated as "weak," as all relevant SNAP measures were not shown to be valid or reliable (either via the use of an objective measure or demonstration of acceptable psychometric properties of a

self-reported measure) [27,30,31,33-37], while 3 studies were classified as "strong," as all SNAP measures used were valid and reliable [28,29,32]. In relation to withdrawals and dropouts, 4 studies were rated as "strong," with retention rates ≥80% (105/121; 935/1071; 367/423; 95/100) [27,33,34,37]; 4 studies were rated as "moderate," with retention rates between 60% and 79% (422/606; 1126/1689; 309/441; 913/1445) [29,30,32,36]; and 3 studies were rated as "weak," with retention rates <60% (148/317; 131/221; 1079/2621) [28,31,35]. In terms of the global rating, 1 study was rated as "moderate" [32], and the remaining 10 studies were rated as "weak" [27-31,33-37].

Table 1. Methodological quality assessment of included studies.

Study	Selection bias	Study design	Confounders	Blinding	Data collection method	Withdrawals and dropouts	Global rating
Nutrition and physical activity studies							
Drieling et al [27]	Weak	Strong	Strong	Weak	Weak	Strong	Weak
Duncan et al [28]	Weak	Strong	Weak	Weak	Strong	Weak	Weak
Franko et al [29]	Weak	Strong	Strong	Weak	Strong	Moderate	Weak
Greene et al [30]	Weak	Strong	Weak	Weak	Weak	Moderate	Weak
McConnon et al [31]	Weak	Strong	Strong	Weak	Weak	Weak	Weak
Patrick et al [32]	Weak	Strong	Strong	Moderate	Strong	Moderate	Moderate
Winett et al [33]	Weak	Strong	Moderate	Weak	Weak	Strong	Weak
Smoking, nutrition, and physical activity study							
Hughes et al [34]	Weak	Strong	Strong	Weak	Weak	Strong	Weak
Smoking, nutrition, alcohol, and physical activity studies							
Cameron et al [35]	Weak	Strong	Strong	Weak	Weak	Weak	Weak
Epton et al [36]	Weak	Strong	Strong	Weak	Weak	Moderate	Weak
Tapper et al [37]	Weak	Strong	Strong	Moderate	Weak	Strong	Weak

Discussion

Principal Findings

This is the first systematic review to examine the long-term effectiveness of internet-based interventions on SNAP behaviors collectively in adults aged 18 years or older. This systematic review focused on internet-based interventions to increase the homogeneity of included studies. This is similar to other systematic reviews that have focused on a specific digital technology [38,39]. More broadly, however, digital technologies can also include text messaging, email, mobile applications, video conferencing [40], and just-in-time feedback interventions [41]. The studies included in this systematic review most commonly examined effectiveness on 2 SNAP behaviors, namely nutrition and physical activity [27-33]. Only 3 studies examined the effectiveness of internet-based interventions on all 4 SNAP behaviors [35-37], whereas 1 study measured the effect on 3 behaviors (ie, tobacco smoking, nutrition, physical activity) [34].

The robust variance estimation meta-analysis findings reported that internet-based interventions achieved an overall significant improvement across all SNAP behaviors in the long term. This suggests that internet-based interventions that adopt a holistic approach to behavior change by addressing multiple SNAP behaviors improve these behaviors collectively and consequently may lead to better health outcomes and reduced health care costs. Given no previous systematic reviews have examined the long-term effectiveness of internet-based interventions on multiple SNAP behaviors nor provided an overall effect size across all SNAP behaviors, we cannot compare our findings to previous reviews. To advance the field, further research is needed on the long-term effectiveness of internet-based interventions on multiple SNAP behaviors.

The studies in this systematic review recruited participants from a variety of settings, including universities [29,30,34-36], the community [28,32,33,37], and health care [27,31]. Among the 5 studies conducted in universities, only 1 study reported a significant treatment effect on both nutrition and physical activity [30], while another study found the intervention reduced

current smoking but had no effect on nutrition, alcohol, and physical activity [36]. Of 4 studies conducted in the community, 2 studies reported a significant treatment effect of the internet-based intervention on nutrition and some physical activity outcomes [32] or on some nutrition outcomes but not physical activity [33]. The trials that recruited participants from health care settings found no significant differences between the internet-based intervention and the control condition [27,31]. Additional research is needed across a variety of settings to expand the evidence base examining the long-term effectiveness of internet-based interventions on combinations of 2 or more SNAP behaviors.

In terms of methodological quality, 10 of 11 studies had a global rating of "weak" [27-31,33-37], with only 1 study rated as "moderate" [32]. Improvements to methodological rigor are particularly needed for selection bias, blinding, and data collection methods. Selection bias could be reduced by using recruitment methods that aim to enroll a representative sample (eg, random selection of potential participants) while data collection methods could be improved via objective measures (eg, pedometers for physical activity, biochemical validation for smoking cessation) for all SNAP behaviors assessed. Given the nature of behavioral interventions, blinding is often difficult; however, future studies should attempt to blind assessors and participants where possible.

Limitations

This systematic review had some limitations. First, although we were able to pool the studies to undertake a robust variance estimation meta-analysis to examine the long-term effectiveness of internet-based interventions across all SNAP behaviors, additional analyses examining potential moderators (eg, country) were not possible due to the relatively small number of studies in this systematic review. Second, the methodological quality assessments were based on the information contained in the articles, and missing details from these articles may have had an impact on the ratings. Finally, all the studies were conducted in high-income countries, which may limit the generalizability of this systematic review's findings to low- and middle-income countries. In addition to expanding the research in the settings and populations included in this review, future research should assess the long-term effectiveness of internet-based interventions on multiple SNAP behaviors in additional populations (eg, culturally and linguistically diverse groups, Indigenous), settings (eg, vocational education settings, rural and remote locations), and countries (eg, low- and middle-income) to strengthen the evidence base and improve the generalizability of the findings.

Conclusions

Internet-based interventions were found to produce an overall significant improvement across all SNAP behaviors in the long term. Given the promising findings on the long-term effectiveness of internet-based interventions across all SNAP behaviors collectively, such interventions may maximize improvements to health and prevent chronic diseases.

Acknowledgments

The School of Medicine and Public Health at the University of Newcastle, Hunter New England Population Health, and the Hunter Medical Research Institute provided infrastructure support. FT was supported by a Heart Foundation Postdoctoral Fellowship (100128), followed by a National Health and Medical Research Council (NHMRC) Career Development Fellowship (APP1143269). The sponsors had no role in study design; in the collection, analyses, and interpretation of data; in the writing of the manuscript; and in the decision to submit the article for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of studies examining the effectiveness of an internet-based intervention on nutrition and physical activity.
[DOCX File , 34 KB - jmir_v23i12e23513_app1.docx]

Multimedia Appendix 2
Characteristics of studies examining the effectiveness of an internet-based intervention on smoking, nutrition, and physical activity.
[DOCX File , 22 KB - jmir_v23i12e23513_app2.docx]

Multimedia Appendix 3
Characteristics of studies examining the effectiveness of an internet-based intervention on smoking, nutrition, alcohol, physical activity (SNAP).
[DOCX File , 26 KB - jmir_v23i12e23513_app3.docx]

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. Khaw K, Wareham N, Bingham S, Welch A, Luben R, Day N. Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. PLoS Med 2008 Jan 08;5(1):e12 [FREE Full text] [doi: 10.1371/journal.pmed.0050012] [Medline: 18184033]
3. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med 2012 Mar 14;19(1):1-13 [FREE Full text] [doi: 10.1007/s12529-010-9135-4] [Medline: 21234735]
4. Kang J, Ciecierski CC, Malin EL, Carroll AJ, Gidea M, Craft LL, et al. A latent class analysis of cancer risk behaviors among U.S. college students. Prev Med 2014 Jul;64:121-125 [FREE Full text] [doi: 10.1016/j.ypmed.2014.03.023] [Medline: 24704131]
5. Paul C, Tzelepis F, Bisquera A, Noble N, Wiggers J. Just how high-risk are ongoing smokers? Exploring clusters of health risk behaviours among current and ex-smokers. Prev Med 2016 Dec;93:70-75. [doi: 10.1016/j.ypmed.2016.09.021] [Medline: 27670372]
6. James E, Freund M, Booth A, Duncan MJ, Johnson N, Short CE, et al. Comparative efficacy of simultaneous versus sequential multiple health behavior change interventions among adults: A systematic review of randomised trials. Prev Med 2016 Aug;89:211-223. [doi: 10.1016/j.ypmed.2016.06.012] [Medline: 27311332]
7. Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. Prev Med 2008 Mar;46(3):181-188 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.001] [Medline: 18319098]
8. Edington DW, Yen LT, Witting P. The financial impact of changes in personal health practices. J Occup Environ Med 1997 Nov;39(11):1037-1046. [doi: 10.1097/00043764-199711000-00004] [Medline: 9383715]
9. Johnson SS, Paiva AL, Cummins CO, Johnson JL, Dyment SJ, Wright JA, et al. Transtheoretical model-based multiple behavior intervention for weight management: effectiveness on a population basis. Prev Med 2008 Mar;46(3):238-246 [FREE Full text] [doi: 10.1016/j.ypmed.2007.09.010] [Medline: 18055007]
10. Bommelaer de Leusse C, Gahnberg C. The Global Internet Report: Consolidation in the Internet Economy. Internet Society. 2019. URL: https://future.internetsociety.org/2019/wp-content/uploads/sites/2/2019/04/InternetSociety-GlobalInternetReport-ConsolidationintheInternetEconomy.pdf [accessed 2020-04-30]
11. Evers KE, Prochaska JM, Prochaska JO, Driskell M, Cummins CO, Velicer WF. Strengths and weaknesses of health behavior change programs on the internet. J Health Psychol 2003 Jan 01;8(1):63-70. [doi: 10.1177/1359105303008001435] [Medline: 22113901]
12. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
13. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013 Sep;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
14. 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]
15. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012 Aug 14;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
16. Meader N, King K, Wright K, Graham HM, Petticrew M, Power C, et al. Multiple risk behavior interventions: meta-analyses of RCTs. Am J Prev Med 2017 Jul;53(1):e19-e30 [FREE Full text] [doi: 10.1016/j.amepre.2017.01.032] [Medline: 28258777]
17. Schoeppe S, Alley S, Van Lippevelde W, 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]
18. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
19. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339(jul21 1):b2535-b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
21. Quality Assessment Tool for Quantitative Studies. Effective Public Health Practice Project (EPHPP). URL: https://merst.ca/wp-content/uploads/2018/02/quality-assessment-tool_2010.pdf [accessed 2020-02-03]

22. Thomas B, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews on Evidence-Based Nursing 2004 Jun 23;1(3):176-184. [doi: 10.1111/j.1524-475x.2004.04006.x]
23. 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]
24. Quality Asesment Tool for Quantitative Studies Dictionary. Effective Public Health Practice Project (EPHPP). URL: https://merst.ca/wp-content/uploads/2018/02/qualilty-assessment-dictionary_2017.pdf [accessed 2020-02-03]
25. Borenstein MH, Hedges LV, Higgins JPT, Rothstein HR. Converting among effect sizes. In: Borenstein MH, Hedges LV, Higgins JPT, Rothstein HR, editors. Introduction to meta-analysis. Hoboken, NJ: John Wiley & Sons; 2009.
26. Higgins JT, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. Cochrane. 2021. URL: https://training.cochrane.org/handbook [accessed 2021-04-30]
27. Drieling RL, Ma J, Thiyagarajan S, Stafford RS. An Internet-based osteoporotic fracture risk program: effect on knowledge, attitudes, and behaviors. J Womens Health (Larchmt) 2011 Dec;20(12):1895-1907. [doi: 10.1089/jwh.2010.2515] [Medline: 21970565]
28. 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]
29. Franko DL, Cousineau TM, Trant M, Green TC, Rancourt D, Thompson D, et al. Motivation, self-efficacy, physical activity and nutrition in college students: randomized controlled trial of an internet-based education program. Prev Med 2008 Oct;47(4):369-377 [FREE Full text] [doi: 10.1016/j.ypmed.2008.06.013] [Medline: 18639581]
30. Greene GW, White AA, Hoerr SL, Lohse B, Schembre SM, Riebe D, et al. Impact of an online healthful eating and physical activity program for college students. Am J Health Promot 2012 Nov 01;27(2):e47-e58. [doi: 10.4278/ajhp.110606-quan-239]
31. McConnon A, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007 Dec 19;7(1):206 [FREE Full text] [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
32. Patrick K, Calfas KJ, Norman GJ, Rosenberg D, Zabinski MF, Sallis JF, et al. Outcomes of a 12-month web-based intervention for overweight and obese men. Ann Behav Med 2011 Dec 6;42(3):391-401. [doi: 10.1007/s12160-011-9296-7] [Medline: 21822750]
33. Winett RA, Anderson ES, Wojcik JR, Winett SG, Bowden T. Guide to health: Nutrition and physical activity outcomes of a group-randomized trial of an internet-based intervention in churches. Ann. Behav. Med 2007 Sep;33(3):251-261. [doi: 10.1007/bf02879907]
34. Hughes SL, Seymour RB, Campbell RT, Shaw JW, Fabiyi C, Sokas R. Comparison of two health-promotion programs for older workers. Am J Public Health 2011 May;101(5):883-890. [doi: 10.2105/ajph.2010.300082]
35. Cameron D, Epton T, Norman P, Sheeran P, Harris PR, Webb TL, et al. A theory-based online health behaviour intervention for new university students (U@Uni:LifeGuide): results from a repeat randomized controlled trial. Trials 2015 Dec 07;16(1):555 [FREE Full text] [doi: 10.1186/s13063-015-1092-4] [Medline: 26643917]
36. Epton T, Norman P, Dadzie A, Harris PR, Webb TL, Sheeran P, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014 Jun 05;14(1):563 [FREE Full text] [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
37. Tapper K, Jiga-Boy G, Maio GR, Haddock G, Lewis M. Development and preliminary evaluation of an internet-based healthy eating program: randomized controlled trial. J Med Internet Res 2014 Oct 10;16(10):e231 [FREE Full text] [doi: 10.2196/jmir.3534] [Medline: 25305376]
38. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
39. Tzelepis F, Paul C, Williams C, Gilligan C, Regan T, Daly J, et al. Real-time video counselling for smoking cessation. Cochrane Database Syst Rev 2019 Oct 29;2019(10):CD012659 [FREE Full text] [doi: 10.1002/14651858.CD012659.pub2] [Medline: 31684699]
40. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HWW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec 4;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
41. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]

Abbreviations

MET: metabolic equivalent of task
NHMRC: National Health and Medical Research Council
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
SNAP: smoking, nutrition, alcohol, physical activity

Edited by G Eysenbach; submitted 14.08.20; peer-reviewed by T Webb, E James; comments to author 21.09.20; revised version received 09.11.20; accepted 22.11.21; published 21.12.21.

Please cite as:
Tzelepis F, Mitchell A, Wilson L, Byrnes E, Haschek A, Leigh L, Oldmeadow C
The Long-Term Effectiveness of Internet-Based Interventions on Multiple Health Risk Behaviors: Systematic Review and Robust Variance Estimation Meta-analysis
J Med Internet Res 2021;23(12):e23513
URL: https://www.jmir.org/2021/12/e23513
doi:10.2196/23513
PMID:34931990

Review

Improving User Experience of Virtual Health Assistants: Scoping Review

Rachel G Curtis[1], BPsych (Hons), PhD; Bethany Bartel[1], BA; Ty Ferguson[1], BPhysio (Hons); Henry T Blake[1], BHmMvt, BHlthSc (Hons); Celine Northcott[1], BBus, MRsch; Rosa Virgara[1], BPhysio (Hons); Carol A Maher[1], BPhysio (Hons), PhD

UniSA Allied Health and Human Performance, Alliance for Research in Exercise, Nutrition and Activity, University of South Australia, Adelaide, Australia

Corresponding Author:
Rachel G Curtis, BPsych (Hons), PhD
UniSA Allied Health and Human Performance
Alliance for Research in Exercise, Nutrition and Activity
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 8 8302 2455
Email: Rachel.Curtis@unisa.edu.au

Abstract

Background: Virtual assistants can be used to deliver innovative health programs that provide appealing, personalized, and convenient health advice and support at scale and low cost. Design characteristics that influence the look and feel of the virtual assistant, such as visual appearance or language features, may significantly influence users' experience and engagement with the assistant.

Objective: This scoping review aims to provide an overview of the experimental research examining how design characteristics of virtual health assistants affect user experience, summarize research findings of experimental research examining how design characteristics of virtual health assistants affect user experience, and provide recommendations for the design of virtual health assistants if sufficient evidence exists.

Methods: We searched 5 electronic databases (Web of Science, MEDLINE, Embase, PsycINFO, and ACM Digital Library) to identify the studies that used an experimental design to compare the effects of design characteristics between 2 or more versions of an interactive virtual health assistant on user experience among adults. Data were synthesized descriptively. Health domains, design characteristics, and outcomes were categorized, and descriptive statistics were used to summarize the body of research. Results for each study were categorized as positive, negative, or no effect, and a matrix of the design characteristics and outcome categories was constructed to summarize the findings.

Results: The database searches identified 6879 articles after the removal of duplicates. We included 48 articles representing 45 unique studies in the review. The most common health domains were mental health and physical activity. Studies most commonly examined design characteristics in the categories of visual design or conversational style and relational behavior and assessed outcomes in the categories of personality, satisfaction, relationship, or use intention. Over half of the design characteristics were examined by only 1 study. Results suggest that empathy and relational behavior and self-disclosure are related to more positive user experience. Results also suggest that if a human-like avatar is used, realistic rendering and medical attire may potentially be related to more positive user experience; however, more research is needed to confirm this.

Conclusions: There is a growing body of scientific evidence examining the impact of virtual health assistants' design characteristics on user experience. Taken together, data suggest that the look and feel of a virtual health assistant does affect user experience. Virtual health assistants that show empathy, display nonverbal relational behaviors, and disclose personal information about themselves achieve better user experience. At present, the evidence base is broad, and the studies are typically small in scale and highly heterogeneous. Further research, particularly using longitudinal research designs with repeated user interactions, is needed to inform the optimal design of virtual health assistants.

(J Med Internet Res 2021;23(12):e31737) doi:10.2196/31737

KEYWORDS

virtual assistant; conversational agent; chatbot; eHealth; digital health; design; user experience; mobile phone

Introduction

Background

Advancements in machine learning and artificial intelligence offer promise for delivering automated, tailored, convenient health assistance with an unprecedented level of sophistication and personalization and are already contributing to the transformation of health care [1]. Virtual assistants can be broadly defined as digital services designed to simulate human conversation and provide personalized responses based on input from the user. They can be programmed with structured conversations or to answer the user's questions. Capabilities range from simple menu or multiple choice–based assistants to more sophisticated virtual assistants with natural language processing that recognize free speech or text. At present, virtual assistants are widely deployed in web-based banking and service settings, reducing reliance on staff by being available to answer consumers' questions about products and services *on demand*. Virtual assistants are also increasingly being designed for various health applications, such as delivering cognitive behavior therapy for depression and anxiety [2], improving diet and physical activity [3], and conducting remote patient monitoring [4]. Despite the exciting potential for using virtual assistants for health purposes, the use of virtual assistants in health could be ineffective or even have unintended negative consequences if the technology does not meet the user's needs and preferences.

The user experience of a virtual health assistant can be defined as the user's perceptions and responses (eg, emotions, beliefs, preferences, and behaviors) that result from its use or anticipated use [5]. User experience is influenced by a range of factors, including presentation, functionality, and interactive behavior [5]. It is important to optimize the design of virtual assistants to provide a positive user experience and promote engagement. A growing body of evidence suggests that design characteristics that influence the *look and feel* of the virtual assistant, such as visual appearance, communication method, and language features, are an important consideration for design, as such design characteristics can significantly influence users' psychological and emotional responses and engagement with technology-based applications [6,7]. In addition, although some design decisions may not affect the cost (eg, whether an avatar should be male or female), other decisions may have a major impact on the cost of designing a virtual health assistant (eg, whether an avatar should be animated with facial expressions). Understanding how such design characteristics influence user experience will assist in using finite health software development budgets most effectively.

Previous literature has proposed general guidelines for designing voice user interfaces [8] and accessible conversational user interfaces for different disability groups [9], as well as virtual assistants for specific purposes such as teaching [10] and in-vehicle assistance [11]. Optimal design techniques are likely to depend on the purpose of the virtual assistant [12,13]; therefore, recommendations specifically in the context of health are needed. Although research has examined methods of assessing the usability of virtual assistants in the health domain [14], clear guidelines on maximizing the user experience of virtual health assistants are lacking.

An important first step toward constructing guidelines for the development of virtual health assistants was achieved by the literature review conducted by ter Stal et al [15] in 2018, which aimed to identify the researched design characteristics for embodied conversational agents (virtual assistants that have an animated avatar) in health. The review provided a comprehensive overview of the existing literature, with results suggesting that speech and/or textual output and facial and gaze expressions were the most commonly researched design characteristics. The secondary aims of ter Stal et al [15] were to identify the outcome variables used in the research and the effects of the design characteristics. The authors concluded that, based on the immature body of evidence at the time, there was no consensus on the optimal design characteristics for embodied conversational agents in health. Results highlighted key avenues for future research, including the fact that more research is needed on all design characteristics to advance the field. Notably, the review by ter Stal et al [15] included studies using any research design and studies where participants viewed stimuli but did not necessarily interact with a virtual assistant.

Objectives

The evidence base for the use of interactive virtual health assistants is rapidly growing in both size and quality. In particular, experimental research designs with interactive virtual assistants are being reported increasingly, which should provide clearer evidence of the influence of design characteristics on user experience. A scoping review methodology offers an explicit, systematic means to overview this large and diverse body of literature using rigorous methods to minimize bias [16]. In this study, we seek to undertake the first scoping review of design characteristics of virtual health assistants, with a view to bring together the strongest evidence available regarding the effects of design characteristics on the user experience of interactive virtual health assistants. In particular, the aims of our scoping review are as follows:

1. Provide an overview of all the experimental research examining how design characteristics of virtual health assistants affect user experience
2. Summarize research findings of experimental research examining how design characteristics of virtual health assistants affect user experience
3. Identify whether research supports making recommendations for the design of virtual health assistants

Bringing together the available evidence on how design characteristics affect the user experience of virtual health assistants will assist researchers and software developers in making decisions about the look and feel of their software and developing the most user-friendly and effective virtual health assistants.

Methods

This review is reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [17].

Eligibility Criteria

Eligibility criteria were designed using the population, intervention, comparator, and outcome framework (population: adults; intervention: virtual health assistant; comparator: design characteristics; and outcome: user experience) [18]. Original research articles in peer-reviewed journals and full-length conference papers were included.

Population

Studies with adult samples (aged ≥18 years) were included.

Intervention

Studies examining virtual health assistants were included. For this review, we considered virtual health assistants to be any virtual assistant aimed at the health consumer (general population or patient) relating to the prevention, management, or treatment of any physical or mental health condition, as well as clinical research. Virtual health assistants were included if they functioned on any electronic device (eg, smartphone, computer, and headset). *Wizard of Oz* virtual assistants (where the user believes they are interacting with a computer-automated virtual assistant, but the virtual assistant is operated by a human [19]) were included.

Comparator

Studies comparing design characteristics between ≥2 versions of a virtual health assistant were included. For this review, we defined design characteristics as characteristics of the virtual assistant that influence its *look and feel* without affecting its core content, purpose, or function. Examples of design characteristics include visual cues such as whether the virtual health assistant has an avatar (ie, an image that represents the virtual assistant), language style, and interaction modality (ie, text or speech). Between- and within-subject experimental designs were included.

Outcome

Studies evaluating user experience outcomes were included. For this review, we defined user experience to include self-reported evaluations of the virtual assistant or the user's interaction with the virtual assistant that indicated a more positive or negative experience (eg, trustworthiness, likeability, enjoyment, and ease of use), affect, intentions to continue using the virtual assistant, and objective measures of user engagement (eg, frequency, duration, or nature of the interaction with the virtual health assistant). Only quantitative data were included.

Exclusion Criteria

Dissertations, review articles, conference abstracts, and studies with children were excluded. Virtual assistants used for training or educating medical professionals, as well as robots with a physical body, were excluded. Studies were excluded if participants did not interact with the virtual health assistant; that is, they did not provide any input into the system. Studies were also excluded if the virtual health assistant was not the main component of the health program. Studies were excluded if they evaluated only 1 version of a virtual assistant (ie, nonexperimental research design with no comparator) or if they compared a virtual assistant to a human. Dependent variables that were not associated with a more positive or negative user experience—for example, those used as *manipulation checks* (eg, where participants were asked to confirm whether a realistic-looking assistant was indeed more realistic looking than a cartoon-style assistant)—were excluded.

Information Sources and Search Strategy

A cross-disciplinary search of the literature was conducted on June 4, 2020, and included 5 electronic databases across the fields of health and information technology: Web of Science, MEDLINE, Embase, PsycINFO, and ACM Digital Library. Search terms for virtual assistant AND design characteristics were included in the search strategy (Table 1). Eligibility specifying the virtual assistant related to *health*, user experience outcomes, and experimental study design was assessed at screening. Searches were limited to the English language with no limit on publication date. Reference lists of the included studies and other key papers in the field were searched to identify further studies (pearling).

Table 1. Search terms.

Search category	Search terms
Virtual assistant	*"conversational agent*"* OR *"conversational system*"* OR *"dialog system*"* OR *"dialogue system*"* OR *"assistance technolog*"* OR *"relational agent*"* OR *"virtual agent*"* OR *"virtual assistant*"* OR *"embodied agent*"* OR *chatbot**
Design characteristics	*anthropomorphi** OR *humanness* OR *personality* OR *emotion** OR *empathy* OR *sympathy* OR *humour* OR *humor* OR *language* OR *linguistic** OR *communication* OR *"conversational tone"* OR *voice* OR *speech* OR *avatar* OR *"profile picture"* OR *face* OR *facial* OR *graphic** OR *appearance* OR *"visual design"* OR *animation* OR *interface* OR *button** OR *menu** OR *emoji** OR *emoticon** OR *"human factors"*

Evidence Selection and Data Charting

Search results from each database were imported into EndNote (Clarivate) [20], in which duplicates were removed. Studies were screened based on title and abstract. Studies that met the eligibility criteria progressed to full-text screening. The full texts of the studies were then screened to determine final eligibility. Articles were screened by 1 of 2 raters. Raters screened a randomly generated selection of 20 articles in duplicate, and the agreement was 100%. A custom form was developed and used for data charting (Multimedia Appendix 1). Extracted data included population, sample size, age, gender, study country, cultural background, health domain, purpose of the virtual assistant, name of the virtual assistant, *Wizard of Oz*

design, device used, animated character, output modality, input modality, whether the interaction was scripted (whether participants were told what to say), duration of interaction, experimental design, and study results. If articles included multiple studies, data extraction was completed only for studies meeting the eligibility criteria. Where multiple eligible studies were included in an article, data were extracted separately. Where relevant outcomes were measured but not compared statistically between experimental conditions, authors were contacted to provide additional information.

Data Synthesis

Study characteristics were compiled for all the studies included in the review. Where a study was reported in multiple articles, articles were compiled as 1 study with a primary reference indicated, as well as an indication of additional references. To facilitate data synthesis across diverse research designs, overarching categories were constructed to describe the health domains, design characteristics, and outcomes. Retrospective thematic analysis was used to identify similar health domains, design characteristics, and outcomes to construct the relevant categories. After data extraction was completed, lists of all reported health domains, design characteristics, and outcomes were compiled. After familiarization with the data, the first author sorted them into similar categories using an inductive approach (ie, directed by the data with no preconceived categories). These categories were reviewed with the senior author, refined, and named.

Data were synthesized descriptively. Descriptive statistics were used to summarize the body of research. A matrix of the design characteristics and outcome categories was constructed to summarize the research findings. Results in the matrix were based on statistical results reported in the articles. Where interactions were examined (eg, in factorial designs or examining interactions with participant characteristics), main effects were included in the matrix. Studies could report results for 1 or multiple outcomes within a particular outcome category. Results were categorized as positive, negative, or no effect. Where studies reported multiple results in a single outcome category, they were categorized as positive if all multiple outcomes showed positive effects, mixed positive if multiple outcomes were reported with both positive and nonsignificant effects, negative if all multiple outcomes showed negative effects, mixed negative if multiple outcomes were reported with both negative and nonsignificant effects, and no effect if multiple outcomes showed no significant effects.

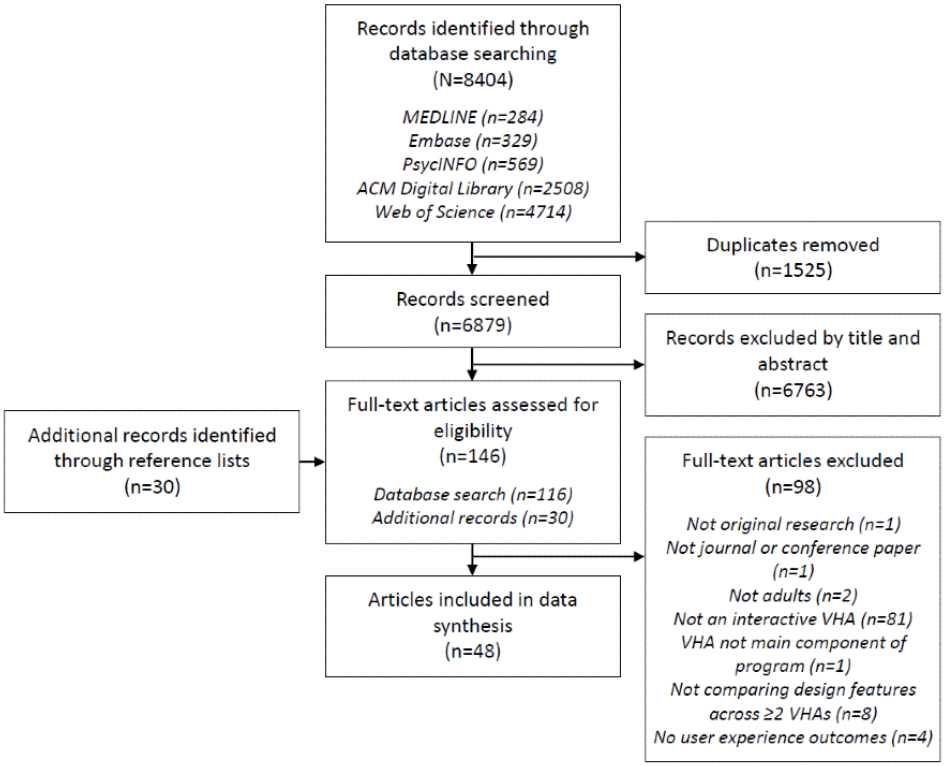
Authors from 2 studies provided additional data on measures that were not compared between experimental groups. Independent sample *t* tests (2-tailed) were conducted, and the results were included in the matrix. In total, 4 studies did not present a statistical analysis comparing relevant experimental conditions; therefore, these studies are included only in the text description.

Results

Overview

The search identified 6879 articles after duplicates were removed. Of the 6879 articles, 6763 (98.31%) were deemed ineligible based on title and abstract screening. We identified 30 additional records through reference lists. In total, 146 articles (116/6879, 1.69% from the database search plus 30 from reference lists) were screened at full text. Of the 146 articles, 98 (67.1%) were deemed ineligible; 81 (55.5%) did not examine an interactive virtual health assistant, 8 (5.5%) did not compare design features between ≥2 virtual health assistants, 4 (2.7%) did not report user experience outcomes, 2 (1.4%) were not adult samples, 1 (0.7%) did not report original research, 1 (0.7%) was not a journal of conference paper, and 1 (0.7%) did not have the virtual health assistant as a main component of the program. Of the 146 articles, a final 48 (32.9%) articles were included in the scoping review (Figure 1). From the 48 articles, 45 unique studies were identified (5 studies were reported in multiple articles, whereas 3 articles contained multiple studies).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. VHA: virtual health assistant.



Multimedia Appendix 2 [21-68] provides an overview of the participant characteristics and study designs for all studies included in the review. Table 2 summarizes the characteristics of the body of research. The virtual assistants used in the research were categorized into 8 health domains: physical activity (aimed to increase exercise), nutrition (aimed to improve diet), alcohol consumption (aimed to reduce alcohol consumption), mental health (eg, aimed to improve mood), medical information or treatment (eg, discussed colorectal cancer screening), sexual health (eg, provided advice about sexually transmitted infections), multiple health behaviors (eg, aimed to improve both exercise and diet), and other (eg, aimed to prevent carpal tunnel). A total of 27 design characteristics were examined in the literature. These were categorized into 5 categories: visual design (eg, realism, age, and body shape of an animated avatar), interface design (eg, input modality), conversational style and relational behavior (eg, empathy and relational behavior and personality), combined visual and conversational design (eg, variability in language and background scene assessed simultaneously), and cultural and organizational affiliation (eg, culturally tailored argumentation and appearance; see Table 3 for the full list of design characteristics). We identified 140 outcome variables, which were categorized into nine categories: virtual assistant personality traits (eg, credible and intelligent), relationship (eg, intimacy and relationship closeness), ease of use (eg, cognitive load and ease of use), satisfaction (eg, enjoyment, satisfaction, and usefulness), emotion (eg, positive and negative affect), use intention (eg, intention to keep using the virtual assistant), engagement (eg, interaction duration), and disclosure (eg, self-disclosure detail and intimacy; see Multimedia Appendix 3 for a full list of outcomes by category). Outcome assessment most frequently used Likert scales, with server logs and conversation transcripts used to assess engagement and disclosure.

Most studies were conducted in the United States, with a greater number of studies conducted during more recent years (ie, between 2017 and 2020). Several authors led multiple studies (ie, Bickmore [25-27,29,31], Creed [34,35], Olaffsson, [53,54], Ring [56-58], and Zhou [67,68]). Most studies examined conversational style and relational behavior or visual design and assessed outcomes in the categories of personality, satisfaction, relationship, and use intention. Virtual assistants most frequently related to mental health and physical activity. Those addressing multiple health behaviors frequently examined physical health and nutrition together. Most virtual assistants had an animated avatar and used speech output and multiple-choice input. Most virtual assistants were automated (did not use a *WizardofOz* design), and participant input was not scripted. Studies were most frequently conducted with between 21 and 100 participants in a single session using a between-subjects design, where participants were allocated to evaluate 1 version of the virtual assistant. Participants were most frequently from the general population, with a larger proportion of females than males. Studies were most often published in conference proceedings in fields related to interdisciplinary research on intelligent virtual agents and human–computer interactions, with fewer published in health-related fields.

Table 2. Summary of study characteristics (N=45).

Study Characteristics	Value, n (%)
Year	
2017-2020	18 (40)
2013-2016	11 (24)
2009-2012	10 (22)
2005-2008	6 (13)
Country	
United States	25 (56)
United Kingdom	4 (9)
Other	7 (16)
Not available	9 (20)
Sample size	
1-20	6 (13)
21-50	16 (36)
51-100	14 (31)
101-200	5 (11)
201-500	4 (9)
Duration	
Single session	37 (82)
Multiple sessions	8 (18)
Health domain	
Mental health	11 (24)
Physical activity	10 (22)
Multiple health behaviors	7 (16)
Medical information or treatment	6 (13)
Nutrition	4 (9)
Sexual health	2 (4)
Alcohol consumption	1 (2)
Other	4 (9)
Design category[a]	
Conversational style and relational behavior	22 (49)
Visual design	12 (27)
Interface design	6 (13)
Cultural and organizational affiliation	5 (11)
Combined visual and conversational design	2 (4)
Outcome category	
Personality	30 (67)
Satisfaction	20 (44)
Relationship	19 (42)
Use intention	17 (38)
Engagement	11 (24)
Ease of use	8 (18)
Emotion	7 (16)

Study Characteristics	Value, n (%)
Disclosure	5 (11)
Virtual assistant characteristics	
Animated avatar	
Yes[b]	31 (69)
No	14 (31)
Output	
Speech[c]	33 (73)
Text	11 (24)
Not available	1 (2)
Input	
Multiple choice	26 (58)
Speech[c]	11 (24)
Text	7 (16)
Not available	1 (2)
Wizard of Oz	
No	37 (82)
Yes	8 (18)
Scripted	
No	41 (91)
Yes	4 (9)

[a]N sums to >45 studies and 100% because 2 studies examined design characteristics in multiple categories.

[b]Includes studies where at least one experimental condition used an animated avatar.

[c]Includes studies where at least one experimental condition used speech.

Table 3. Summary of research findings (N=41).

Design characteristics	Values, n (%)	Outcomes (effect)							
		Personality	Relationship	Ease of use	Satisfaction	Emotion	Use intention	Engagement	Disclosure
Visual design									
Animated avatar (vs no visual representation)	3 (7)	• Mixed negative [48][a],[62] • No significant effect [51]	Mixed negative [48][a],[62]	No significant effect [48][a]	No significant effect [48][a],[51,62]	No significant effect [48][a],[51]	No significant effect [48][a]	—[b]	—
Realistic (vs cartoon)	4 (10)	• Positive [64] • Mixed negative [57] • No significant effect [57,62]	No significant effect [62]	No significant effect [64]	• Mixed positive [64] • No significant effect [62]	—	• Positive [64] • No significant effect [57][c] (2 studies)	—	—
Human (vs robot)	1 (2)	Mixed negative [62]	Mixed negative [62]	—	Mixed negative [62]	—	—	—	—
Younger (vs older)	1 (2)	No significant effect [64]	—	No significant effect [64]	Mixed positive [64]	—	No significant effect [64]	—	—
Fat (vs slim)	1 (2)	Positive [63]	No significant effect [63]	—	—	—	No significant effect [63]	—	—
Familiar (vs unfamiliar)	1 (2)	Mixed negative [64]	—	No significant effect [64]	Mixed negative [64]	—	No significant effect [64]	—	—
Medical professional attire (vs casual)	1 (2)	Positive [55]	Positive [55]	—	—	—	Positive [55]	—	—
Medical office (vs empty room)	1 (2)	Mixed positive [55]	No significant effect [55]	—	—	—	No significant effect [55]	—	—
Variability in camera angle (vs no variability)	3 (7)	• Mixed positive [58] • No significant effect [58][c] (2 studies)	—	—	—	—	—	No significant effect [58][c] (3 studies)	—
Interface design									
Speech input (vs text or multiple choice)	3 (7)	No significant effect [33]	No significant effect [33]	Mixed negative [32]	—	—	—	Positive [52]	—
Motion initiated (vs user initiated)	1 (2)	—	Positive [56]	—	—	Positive [56]	—	No significant effect [56]	—

Design characteristics	Values, n (%)	Outcomes (effect)							
		Personality	Relationship	Ease of use	Satisfaction	Emotion	Use intention	Engagement	Disclosure
Polite notification ringtone (vs impolite)	1 (2)	Positive [25]	—	—	—	—	Positive [25]	—	—
Conversational style and relational behavior									
Empathy and relational behavior (vs none)	7 (17)	• Positive [48,51] • Mixed positive [31] • No significant effect [40][d],[49]	• Mixed positive [48] • No significant effect [27,49]	No significant effect [48]	• Positive [48] • Mixed positive [27,31,51] • No significant effect [49]	• Mixed positive [27,39] • No significant effect [48,51]	• Positive [31,48] • No significant effect [27]	No significant effect [40]	Positive [31]
Emotional expression (vs none)	3 (7)	Positive [34,43]	No significant effect [34,35,43]	—	No significant effect [43]	No significant effect [34]	—	—	—
Self-disclosure (vs none)	3 (7)	• Positive [47][e] • No significant effect [29]	Positive [44,47]	—	• Positive [47] • Mixed positive [29]	—	—	Positive [29,47]	• Mixed positive [47][e] • Positive [44]
Personality (various)[f]	3 (7)	• Positive [61] • No significant effect [36]	• Mixed positive [60] • No significant effect [36]	• Positive [61] • No significant effect [36]	No significant effect [36]	No significant effect [36]	No significant effect [36,60]	No significant effect [36]	—
Conversation memory (vs none)	2 (5)	• Mixed positive [36] • No significant effect [23]	Mixed positive [36]	No significant effect [36]	No significant effect [23,36]	No significant effect [36]	No significant effect [23,36]	No significant effect [23,36]	—
Humor (vs none)	1 (2)	Mixed positive [36]	No significant effect [36]	No significant effect [36]	Mixed positive [36]	No significant effect [36]	No significant effect [36]	No significant effect [36]	—
Emojis (vs none)	1 (2)	No significant effect [37]	—	—	No significant effect [37]	—	—	No significant effect [37]	—
Rap (vs none)	1 (2)	Mixed negative [53]	Mixed positive [53]	—	No significant effect [53]	—	No significant effect [53]	—	—
Participant control of facial and vocal expression (vs none)	1 (2)	—	—	No significant effect [26]	Mixed positive [26]	—	—	—	—

Design characteristics	Values, n (%)	Outcomes (effect)							
		Personality	Relationship	Ease of use	Satisfaction	Emotion	Use intention	Engagement	Disclosure
Constrained to positive user response options (vs negative responses allowed)	1 (2)	No significant effect [54]	No significant effect [54]	—	Mixed negative [54]	—	No significant effect [54]	—	—
Combined visual and conversational design									
Personification (name, static avatar, and conversational language vs none)	1 (2)	—	—	—	—	—	—	—	Mixed negative [59]
Variability in dialog structure, language, and scene (vs no variability)	1 (2)	Positive [29]	—	—	—	—	Positive [29]	Positive [29]	—
Cultural and organizational affiliation									
Culturally tailored argumentation (vs not)	2 (5)	No significant effect [65]	—	—	Positive [50]	—	—	—	—
Culturally tailored appearance (vs not)	3 (7)	No significant effect [65,67]	Negative [67]	No significant effect [67]	No significant effect [50,67]	—	No significant effect [67]	—	—
Culturally tailored argumentation and scene combined (vs not)	1 (2)	No significant effect [68]	No significant effect [68]	No significant effect [68]	No significant effect [68]	—	No significant effect [68]	—	—
Patient assistant (vs researcher or government employee)	1 (2)	Mixed positive [66]	—	—	Mixed positive [66]	—	Positive [66]	—	—

[a]Results indicated for nonempathetic avatar only (empathetic avatar had additional dialog to the no avatar condition).

[b]No study examined the combination of design characteristic and outcome.

[c]Multiple studies were reported in the article with similar results.

[d]Similar results were additionally reported at a different time point in the study [39].

[e]Similar results were additionally reported at a different time point in the study [46].

[f]Indicates any effects of personality (no consistent comparator).

Table 3 summarizes research findings grouped according to the design characteristic examined and the categories of outcomes measured. Where identical outcomes of a study were reported in multiple articles, the primary reference listed in Multimedia Appendix 2 was used. Additional references were used for outcomes that were not reported in the primary study. In total, 4 studies did not present a statistical analysis comparing the relevant experimental conditions; therefore, these studies are not included in Table 3.

The following paragraphs highlight key results from the studies presented in Table 3 and include a narrative synthesis of studies that were not presented in Table 3.

Visual Design

Approximately 7% (3/41) of studies examined whether user experience differed using a virtual assistant with an animated avatar compared with using a text- or speech-only virtual assistant with no visual representation [48,51,62]. Findings were generally nonsignificant [48,51,62], with some mixed negative effects of using an animated avatar [48,62]. An additional study not included in Table 3 concluded that virtual assistants with an animated avatar were preferred over voice-only assistants; however, the analyses included both real and virtual assistants [45].

Approximately 22% (9/41) of studies examined the appearance of the animated avatar, and 10% (4/41) of studies examined whether user experience differed using a virtual assistant with a more realistic human avatar compared with a more cartoon human avatar [57,62,64]. Although some positive and mixed positive effects of using a more realistic avatar were found [64], more effects were nonsignificant [57,62,64], and 1 was negative [57]. The species of the avatar was examined by 2% (1/41) of studies, which found mixed negative effects of using a human avatar compared with using a robot avatar [62]. Age was examined by 2% (1/41) of studies, which found mixed positive effects of using an avatar with a younger appearance compared with using one with an older appearance on satisfaction but no significant effects on other outcomes [64]. Body shape was examined by 2% (1/41) of studies, which found a positive effect of a fat avatar compared with a slim avatar on personality traits but nonsignificant effects on other outcomes [63]. The familiarity of the avatar was examined by 2% (1/41) of studies, which found mixed negative and nonsignificant effects of using an avatar that looked like a health coach that participants met at the beginning of the session compared with using an unfamiliar avatar [64]. The avatar's attire was examined by 2% (1/41) of studies, which found consistently positive effects of medical professional attire compared with casual attire [55].

The background scene behind the avatar was examined by 2% (1/41) of studies, which found mixed positive effects of representing a medical office compared with representing an empty room on personality but no significant effects on other measured outcomes [55]. Approximately 7% (3/41) of studies (all reported in 1 paper) examined whether variability in the *camera* position, distance, and focus was associated with user experience and found mostly nonsignificant effects [58].

Interface Design

Approximately 7% (3/41) of studies examined the effects of input modality—whether the user communicates using speech, text, or multiple choice—on user experience and found a combination of positive, mixed negative, and nonsignificant effects of speech input compared with other modalities [32,33,52]. A menu-based virtual assistant was examined by 1 further study not included in Table 3, and it concluded that there were no differences in usability between speech and phone key press user input [42].

How the conversation between the virtual assistant and user was initiated was examined by 2% (1/41) of studies, which found positive and nonsignificant effects of automated motion initiation compared with user initiation [56]. The type of ringtone used to initiate a conversation with the user was examined by 2% (1/41) of studies, which found positive effects of more polite tones compared with less polite tones [25].

Conversational Style and Relational Behavior

Approximately 17% (7/41) of studies examined empathy and relational behavior—empathetic verbal feedback and nonverbal behavior such as facial expressions and gestures [27,31,39,40,48,49,51]. Although some effects were nonsignificant [27,40,48,49,51], more effects were positive or mixed positive, with 71% (5/7) of studies showing at least some positive effect [27,31,39,48,51]. Approximately 7% (3/41) of studies examined emotional expression—the use of facial expression and voice to express emotion—and found some mixed positive effects [34,43] but more nonsignificant effects [34,35,43]. Approximately 7% (3/41) of studies examined self-disclosure—whether the virtual assistant tells the user information about themselves—and found mostly positive effects [29,44,47]. Approximately 7% (3/41) of studies examined personality [36,60,61]. Although some positive and mixed positive effects were found [60,61], most effects were nonsignificant [36,60].

Approximately 5% (2/41) of studies examined conversation memory—whether the virtual assistant remembered information from earlier conversation—and found some mixed positive effects [36] but mostly nonsignificant effects [23,36]. An additional study not included in Table 3 compared users' first interactions when the virtual assistant did not recall their previous session and when the virtual assistant did recall their previous session [38]. The authors concluded that users were more positive when the virtual assistant recalled their session; however, the conversations were less personal.

Humor was examined by 2% (1/41) of studies, which found mostly nonsignificant effects of including humor compared with not including humor [36]. Using emojis was examined by 2% (1/41) of studies, which found no significant effects of using emojis compared with not using emojis [37]. Rap was examined by 2% (1/41) of studies, which found a combination of mixed positive, mixed negative, and nonsignificant effects of including rap compared with not including rap [53]. Allowing participants to control the virtual assistant's facial and vocal expression was examined by 2% (1/41) of studies, which found nonsignificant and mixed positive effects compared with not allowing such control [26]. Approximately 2% (1/41) of studies examined constraining users to respond only positively to questions about their confidence and motivation compared with also presenting negative multiple-choice response options [54]. It found a combination of mixed negative and neutral effects of constraining users to positive responses. A further study not included in Table 3 examined whether user evaluations were more positive for a virtual assistant that changed behavior based on the user's eye contact compared with a virtual assistant that always appeared attentive or always bored or that changed behavior randomly [41]. The authors concluded that changing based on the user's eye contact seemed more normal than changing behavior randomly but did not confirm the hypothesis

that changing behavior is more normal than unchanging behavior.

Combined Visual and Conversational Design

Personification—the use of a name, static avatar, and conversational language—was examined by 2% (1/41) of studies, which found negative effects of personification on users' disclosure [59]. Variability in dialog structure (the order of the conversation and the utterances used) and background scene was examined by 2% (1/41) of studies, which found consistently positive effects of variability compared with no variability [29].

Organizational and Cultural Affiliation

Approximately 10% (4/41) of studies examined cultural tailoring—matching the culture of the virtual assistant to that of the user [50,65,67,68]. Approximately 5% (2/41) of studies examined cultural tailoring of the virtual assistant's argumentation (eg, discussed culturally relevant topics) [50,65], and 50% (1/2) of those found a positive effect [50]. Approximately 7% (3/41) of studies examined cultural tailoring of the virtual assistant's appearance and the household setting and found predominantly nonsignificant effects [50,65,67]. Culturally tailored background scene and argumentation combined were examined by 2% (1/41) of studies, which found no significant effects [68]. The organizational affiliation of the virtual assistant—who the virtual assistant claimed to be and the context provided in the background scene—was examined by 2% (1/41) of studies, which found positive effects of the virtual assistant being a patient assistant compared with the virtual assistant being either a member of the medical team conducting the research or a government employee [66].

Discussion

Principal Findings

This study aimed to provide an overview of experimental research examining how design characteristics of virtual health assistants affect user experience. This is a growing area of scientific endeavor with studies, taken together, examining highly diverse health domains, design characteristics, and outcomes. The most common health domains were physical activity and mental health, with relatively few virtual assistants related to specific health conditions. Approximately half of the studies were categorized as examining the design of conversational style and relational behavior, with the most common design characteristic researched being empathy and relational behavior. The most commonly measured outcomes were in the categories of personality traits, satisfaction, relationship, and use intention.

This study also aimed to summarize the research findings of experimental research examining how design characteristics of virtual health assistants affect user experience. Generally, research has been piecemeal, with few design characteristics having a sufficient body of evidence to draw conclusions about their effects on user experience. The 2 design characteristics that defy this are virtual assistants' empathy and relational behavior and self-disclosure, which have been the focus of a good number of studies. Research suggests that all 3 (ie, empathy, relational behavior, and self-disclosure) are related to more positive user experience. Other design characteristics with emerging levels of evidence are having a more realistic human representation for an avatar and having medical attire for the avatar, both of which may potentially be related to more positive user experience. Finally, evidence to date suggests that using an animated avatar (compared with no avatar) and cultural tailoring may not affect user experience; however, more research is needed to explore these findings.

One of the clearest findings of this study was that the use of empathy and relational behavior in virtual health assistants appears to have positive effects on user experience. Empathy may help to build trust and rapport with the virtual assistant. The finding that empathy was associated with user satisfaction is in line with research indicating a positive association between empathy in real health care providers and patient satisfaction [69,70]. Results were not consistently positive; however, this may be related to differences between the virtual assistants. For example, for the outcome category personality traits, of the 5 studies examining empathy and relational behavior, 3 (60%) studies showing positive effects used animated avatars, including nonverbal relational behaviors [31,48,51]. In contrast, 40% (2/5) of studies showing no effects were text-only assistants [40,49]. It may be that users do not expect text-only assistants to show empathy; therefore, the presence or absence of empathy has no impact on the ratings of the virtual assistant. Alternatively, the effects of empathy may be diminished when nonverbal relational behaviors such as expression and gestures are not present.

Research suggests that virtual health assistants that use self-disclosure (ie, provide information about themselves) elicit a more positive user experience. Results were similar whether the autobiographical information was framed as being about the virtual assistant's experience as a computer agent [44] or included human experiences that could not actually be true [29,47]. Self-disclosure is important for the formation of relationships [71], although research suggests that self-disclosure by a real counselor can have either positive or detrimental effects on a client's perceptions of the counselor [72]. The finding that users respond positively to the autobiographical stories of a virtual health assistant supports the *computers are social actors* paradigm, where users display social responses to computers, although they know they are not human [73,74].

Research examining the realism of the animated avatar showed some positive effects; however, more were nonsignificant. The *uncanny valley* theory suggests that robots that appear almost but not quite human may elicit a negative emotional response and be less likable than those that are clearly nonhuman [75]. However, in this review, the study that used a photo-realistic representation in the realistic experimental condition [64] showed positive effects. More research is needed to examine how the realism of the avatar affects the user experience of virtual health assistants.

Results from 1 study suggest that dressing the avatar in medical attire results in a more positive user experience [55]. Although more research is needed to confirm this finding, this was a large study (n=308) with consistent results across all outcomes measures. Interestingly, the background setting for the avatar

(medical office or empty room) had a mixed positive effect on only 1 out of 3 outcomes categories [55].

Research suggests that including an animated avatar has no effect or, in some cases, a negative effect on user experience. However, upon closer inspection, this may be because of the nature of the avatars used in the research and may also be affected by interactions between the animation and other virtual assistant characteristics. For example, Lisetti et al [48] showed that an animated avatar with a neutral facial expression and no empathetic dialog led to poorer user experience than a text-only virtual assistant, whereas an expressive and empathetic virtual assistant led to a better user experience than the text-only virtual assistant. Nguyen and Masthoff [51] reported similar findings; a nonempathetic animated virtual assistant and a nonempathetic text-only virtual assistant led to a similar user experience; however, an empathetic animated virtual assistant led to better user experience than an empathetic text-only virtual assistant. Taken together, it appears that users may expect a virtual assistant with a human-like representation to have empathy and human-like relational behaviors and have a poorer user experience when this expectation is not met.

Overall, the research did not show cultural tailoring to improve the user experience of virtual health assistants. Notably, although 75% (3/4) of studies included participants who were born overseas (in China [68], India [50], or a Spanish-speaking Latin-American country [65]), participants in all the studies lived in the United States. This may suggest that cultural tailoring is not required for different cultures living in the United States who have had exposure to Anglo-American culture, although more research could confirm this finding. Additional research is also needed to determine whether cultural tailoring affects user experience in other cultural contexts.

Strengths and Limitations

This scoping review is the most rigorous attempt at synthesizing the literature regarding the effects of design characteristics on the user experience of virtual health assistants. It followed the PRISMA-ScR guidelines for scoping reviews and searched a large number of databases. It examined a broad range of design characteristics using the highest level of evidence—experimental research using only interactive virtual health assistants where participants were able to input into the system. However, we acknowledge that the use of specific search terms to capture virtual assistants and design characteristics could have omitted some results. It is also possible that other literary sources may have been available in other databases. In addition, qualitative data were excluded. This enabled a structured approach to synthesizing the data based on statistical significance but may have omitted some important views on user experience.

Although the breadth of the review is a major strength, the heterogeneity of the included studies makes it difficult to synthesize and interpret the results. There was considerable heterogeneity in the purpose of the virtual assistants studied. Optimal design techniques may differ among different health domains. For example, although no overall effect of using emojis was found, the difference in ratings of confidence between using text-only and text with emojis depended on whether the virtual assistant was discussing physical or mental well-being [37]. In addition, some health conditions were not represented in the studies, for example, neurocognitive impairments such as dementia. There was also significant heterogeneity in the outcomes measured. The most commonly measured outcomes were in the categories of personality, satisfaction, relationship, and use intention. Few studies examined the ease of use, engagement, or disclosure. Although interface design may play a key role in determining the ease of use, other design characteristics such as the visual appearance of an avatar may not be expected to affect the ease of use. More research examining how users interact with the virtual assistant (engagement and disclosure), particularly using objective measures, may complement subjective ratings of the virtual assistant and interaction.

An additional limitation of the literature is that some studies combined a set of similar characteristics into 1 condition, making it difficult to ascertain which characteristic might be responsible for the effects on user experience. For example, research on empathy and relational behavior frequently included verbal empathy with nonverbal relational behaviors. In addition, in most studies, participants evaluated the virtual assistant after interacting during a single session. Programs that aim to promote health behavior change or provide support for a health condition are often designed for ongoing use. Additional research should examine how design characteristics affect user experience over time. Most virtual assistants had animated avatars and speech output; however, over half constrained user input to selecting from predefined response options. Constraining user input requires simpler programming and removes the risk of errors occurring when the virtual assistant misinterprets the user's input or cannot formulate a response to a query that is outside the bounds of its programmed knowledge [76]. Natural language processing enables users to communicate using unconstrained text or speech and enables more natural user-directed communication. Virtual assistants using natural language processing have been commonly used in health care [77] and, with rapid advancements in artificial intelligence, are likely to become increasingly sophisticated. More research should examine the design and user experience of these types of virtual health assistants.

Recommendations

Research demonstrates that design characteristics affect the user experience of virtual health assistants; therefore, researchers and software developers should carefully consider the look and feel of a virtual health assistant during development and testing. On the basis of the results of this scoping review, the following recommendations for designing virtual health assistants and advancing the field of research may be useful for health researchers and software developers:

1. Design virtual health assistants to express verbal empathy, for example, understanding of the user's feelings
2. Design virtual health assistants to disclose personal information about themselves to the user, for example, information about their past and personal preferences
3. Consider designing a human avatar to be more realistic with medical professional attire

4. If designing an animated virtual health assistant, it should display nonverbal relational behaviors, for example, emotional facial expressions, gestures, and mutual gaze
5. If empathy and relational behaviors are unable to be incorporated, consider that an animated avatar may not be beneficial or cost-effective
6. Engage in formative research with the target audience and adopt a user-centered design approach to ensure that the software meets the needs and preferences of the user
7. Conduct further systematic research to replicate and extend previous findings, particularly with longitudinal research designs with repeated user interactions, objective engagement outcomes, and virtual assistants with natural language processing capabilities

Conclusions

Virtual health assistants can provide health information and support on demand and may be applied in the future to a wide variety of purposes such as providing public health information, health education, supporting patients with chronic health conditions, and assisting with healthy lifestyle behavior change. This scoping review examined experimental research assessing how design characteristics of virtual health assistants affect user experience. This is a rapidly growing field of research but is difficult to synthesize and interpret because of the heterogeneity of studies. Nonetheless, certain design characteristics have emerged as important for improving user experience. Preliminary recommendations suggest that programming virtual health assistants to show empathy, display nonverbal relational behaviors, and disclose personal information about themselves may result in a more positive user experience. The decision to include an animated avatar should consider whether the avatar can display empathy and nonverbal relational behaviors. Future research is required to improve our understanding of the relationship between design characteristics and user experience of virtual health assistants, particularly with longitudinal research designs with repeated user interactions.

Acknowledgments

CAM is supported by a Medical Research Future Fund Emerging Leader Grant (GNT1193862). HTB is supported by an Australian Government Research Training Program Scholarship. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

RGC and CAM conceived and designed the study. RGC conducted database searches. RGC and BB conducted screening, and RGC, BB, TF, CN, HTB, and RV conducted data extraction. RGC analyzed the data and wrote the manuscript. All authors revised the manuscript and reviewed and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data charting form.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i12e31737_app1.xlsx]

Multimedia Appendix 2
Characteristics of studies included in the scoping review.
[DOCX File , 56 KB - jmir_v23i12e31737_app2.docx]

Multimedia Appendix 3
Outcome categories.
[DOCX File , 17 KB - jmir_v23i12e31737_app3.docx]

References

1. Horgan D, Romao M, Morré SA, Kalra D. Artificial intelligence: power for civilisation - and for better healthcare. Public Health Genomics 2019;22(5-6):145-161 [FREE Full text] [doi: 10.1159/000504785] [Medline: 31838476]
2. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
3. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17558 [FREE Full text] [doi: 10.2196/17558] [Medline: 32673246]
4. Cleres D, Rassouli F, Brutsche M, Kowatsch T, Barata F. Lena: a voice-based conversational agent for remote patient monitoring in chronic obstructive pulmonary disease.; April 13-17 2021; College Station, USA. In: Proceedings of the 26th

ACM Conference on Intelligent User Interfaces. 2021 Presented at: 26th ACM Conference on Intelligent User Interfaces; Apr 13-17, 2021; College Station, USA. [doi: 10.3929/ethz-b-000472586]
5. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. ISO. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2021-11-29]
6. Thüring M, Mahlke S. Usability, aesthetics and emotions in human–technology interaction. Int J Psychol 2007 Aug;42(4):253-264. [doi: 10.1080/00207590701396674]
7. Bhandari U, Chang K, Neben T. Understanding the impact of perceived visual aesthetics on user evaluations: an emotional perspective. Inf Manag 2019 Jan;56(1):85-93. [doi: 10.1016/j.im.2018.07.003]
8. Pearl C. Designing Voice User Interfaces: Principles of Conversational Eperiences. Beijing, China: O'Reilly Media Incorporated; 2016.
9. Lister K, Coughlan T, Iniesto F, Freear N, Devine P. Accessible conversational user interfaces: considerations for design. In: Proceedings of the 17th International Web for All Conference. 2020 Presented at: Proceedings of the 17th International Web for All Conference; Apr 20-21 2020; Taipei, Taiwan. [doi: 10.1145/3371300.3383343]
10. Wellnhammer N, Dolata M, Steigler S, Schwabe G. Studying with the help of digital tutors: design aspects of conversational agents that influence the learning process. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; Jan 2020; Hawaii, USA. [doi: 10.24251/hicss.2020.019]
11. Strohmann T, Siemon D, Robra-Bissantz S. Designing virtual in-vehicle assistants: design guidelines for creating a convincing user experience. AIS Trans Human Comput Interac 2019;11(2):54-78. [doi: 10.17705/1thci.00113]
12. Følstad A, Skjuve M, Brandtzaeg P. Different chatbots for different purposes: towards a typology of chatbots to understand interaction design. In: Internet Science. Cham: Springer; 2019.
13. Janssen A, Passlick J, Rodríguez Cardona D, Breitner MH. Virtual assistance in any context. Bus Inf Syst Eng 2020 Apr 06;62(3):211-225. [doi: 10.1007/s12599-020-00644-1]
14. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, Mctear M. Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: Proceedings of the 31st European Conference on Cognitive Ergonomics. 2019 Presented at: 31st European Conference on Cognitive Ergonomics; Sept 10-13, 2019; Belfast, UK. [doi: 10.1145/3335082.3335094]
15. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. Int J Human Comput Stud 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
16. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
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 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123(3):A12-A13. [Medline: 7582737]
19. Dahlbäck N, Jönsson A, Ahrenberg L. Wizard of Oz studies — why and how. Knowl Based Syst 1993 Dec;6(4):258-266. [doi: 10.1016/0950-7051(93)90017-n]
20. Endnote Team. EndNote. EndNote X9 ed. Philadelphia, PA: Clarivate; 2013.
21. Amini R, Lisetti C, Yasavur U. Emotionally responsive virtual counselor for behavior-change health interventions. In: Advancing the Impact of Design Science: Moving from Theory to Practice. Cham: Springer International Publishing; 2014.
22. Amini R, Lisetti C, Yasavur U, Rishe N. On-demand virtual health counselor for delivering behavior-change health interventions. In: Proceedings of the IEEE International Conference on Healthcare Informatics. 2013 Presented at: IEEE International Conference on Healthcare Informatics; Sep 9-11, 2013; Philadelphia, PA, USA. [doi: 10.1109/ichi.2013.13]
23. Bellur S, Sundar SS. Talking health with a machine: how does message interactivity affect attitudes and cognitions? Hum Commun Res 2016 Sep 15;43(1):25-53. [doi: 10.1111/hcre.12094]
24. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
25. Bickmore T, Mauer D, Crespo F, Brown T. Persuasion, task interruption and health regimen adherence. In: Proceedings of the Persuasive Technology, Second International Conference on Persuasive Technology, PERSUASIVE 2007. 2007 Presented at: Persuasive Technology, Second International Conference on Persuasive Technology, PERSUASIVE 2007; Apr 26-27, 2007; Palo Alto, CA, USA. [doi: 10.1007/978-3-540-77006-0_1]
26. Bickmore T, Ring L. Making it personal: end-user authoring of health narratives delivered by virtual agents. In: Proceedings of the Intelligent Virtual Agents, 10th International Conference, IVA 2010. 2010 Presented at: Intelligent Virtual Agents, 10th International Conference, IVA 2010; Sep 20-22, 2010; Philadelphia, PA, USA. [doi: 10.1007/978-3-642-15892-6_43]
27. Bickmore T, Schulman D. Practical approaches to comforting users with relational agents. In: Proceedings of the CHI Conference on Human Factors in Computing. 2007 Presented at: CHI07: CHI Conference on Human Factors in Computing; Apr 28-May 3, 2007; San Jose, CA, USA. [doi: 10.1145/1240866.1240996]

28. Bickmore T, Schulman D, Yin L. Engagement vs. deceit: virtual humans with human autobiographies. In: Proceedings of the Intelligent Virtual Agents, 9th International Conference, IVA 2009. 2009 Presented at: Intelligent Virtual Agents, 9th International Conference, IVA 2009; Sep 14-16, 2009; Amsterdam, The Netherlands. [doi: 10.1007/978-3-642-04380-2_4]
29. Bickmore T, Schulman D, Yin L. Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
30. Bickmore T, Picard R. Towards caring machines. In: Proceedings of the CHI 2004 Conference on Human Factors in Computing. 2004 Presented at: CHI 2004 Conference on Human Factors in Computing; Apr 24-29, 2004; Vienna, Austria. [doi: 10.1145/985921.986097]
31. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun 01;12(2):293-327. [doi: 10.1145/1067860.1067867]
32. Chen J, Lyell D, Laranjo L, Magrabi F. Effect of speech recognition on problem solving and recall in consumer digital health tasks: controlled laboratory experiment. J Med Internet Res 2020 Jun 01;22(6):e14827 [FREE Full text] [doi: 10.2196/14827] [Medline: 32442129]
33. Cho E. Hey Google, can i ask you something in private? In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300488]
34. Creed C, Beale R. User interactions with an affective nutritional coach. Interact Comput 2012 Sep;24(5):339-350. [doi: 10.1016/j.intcom.2012.05.004]
35. Creed C, Beale R, Cowan B. The impact of an embodied agent's emotional expressions over multiple interactions. Interact Comput 2014 Jan 29;27(2):172-188. [doi: 10.1093/iwc/iwt064]
36. De Boni M, Richardson A, Hurling R. Humour, relationship maintenance and personality matching in automated dialogue: a controlled study. Interact Comput 2008 May;20(3):342-353. [doi: 10.1016/j.intcom.2007.11.002]
37. Fadhil A, Schiavo G, Wang Y, Yilma B. The effect of emojis when interacting with conversational interface assisted health coaching system. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: Pervasive Health '18: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21 - 24, 2018; New York NY USA. [doi: 10.1145/3240925.3240965]
38. Ferland L, Koutstaal W. How's your day look? The (un)expected sociolinguistic effects of user modeling in a conversational agent. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020; Honolulu, HI, USA. [doi: 10.1145/3334480.3375227]
39. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. Towards understanding emotional intelligence for behavior change chatbots. In: Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction. 2019 Presented at: 8th International Conference on Affective Computing and Intelligent Interaction; Sep 3-6, 2019; Cambridge, UK. [doi: 10.1109/acii.2019.8925433]
40. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. EMMA: an emotion-aware wellbeing chatbot. In: Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction. 2019 Presented at: 8th International Conference on Affective Computing and Intelligent Interaction; Sep 3-6, 2019; Cambridge, UK. [doi: 10.1109/acii.2019.8925455]
41. Grillon H, Thalmann D. Eye contact as trigger for modification of virtual character behavior. In: Proceedings of the Virtual Rehabilitation. 2008 Presented at: Virtual Rehabilitation; Aug 25-27, 2008; Vancouver, BC, Canada. [doi: 10.1109/icvr.2008.4625161]
42. Grover A, Plauché M, Barnard E, Kuun C. HIV health information access using spoken dialogue systems: touchtone vs. speech. In: Proceedings of the International Conference on Information and Communication Technologies and Development. 2009 Presented at: International Conference on Information and Communication Technologies and Development; Apr 17-19, 2009; Doha, Qatar. [doi: 10.1109/ictd.2009.5426716]
43. Guadagno RE, Swinth KR, Blascovich J. Social evaluations of embodied agents and avatars. Comput Human Behav 2011 Nov;27(6):2380-2385. [doi: 10.1016/j.chb.2011.07.017]
44. Kang S, Gratch J. People like virtual counselors that highly-disclose about themselves. Stud Health Technol Inform 2011;167:143-148. [Medline: 21685657]
45. Kim K, Norouzi N, Losekamp T, Bruder G, Anderson M, Welch G. Effects of patient care assistant embodiment and computer mediation on user experience. In: Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). 2019 Presented at: 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Dec 9-11, 2019; San Diego, CA. [doi: 10.1109/aivr46125.2019.00013]
46. Lee Y, Yamashita N, Huang Y. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. Proc ACM Human Comput Interact 2020 May 28;4(1):1-27. [doi: 10.1145/3392836]
47. Lee Y, Yamashita N, Huang Y, Fu W. "I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020; Honolulu, HI, USA. [doi: 10.1145/3313831.3376175]

48. Lisetti C, Amini R, Yasavur U, Rishe N. I can help you change! An empathic virtual agent delivers behavior change health interventions. ACM Trans Manage Inf Syst 2013 Dec 01;4(4):1-28. [doi: 10.1145/2544103]
49. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
50. Murali P, Shamekhi A, Parmar D, Bickmore T. Argumentation is more important than appearance for designing culturally tailored virtual agents. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. 2020 Presented at: 19th International Conference on Autonomous Agents and MultiAgent Systems; May 9 - 13, 2020; Auckland, New Zealand. [doi: 10.1145/860575.860717]
51. Nguyen H, Masthoff J. Designing empathic computers: the effect of multimodal empathic feedback using animated agent. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: 4th International Conference on Persuasive Technology; Apr 26-29, 2009; Claremont, CA, USA. [doi: 10.1145/1541948.1541958]
52. Novielli N, de Rosis F, Mazzotta I. User attitude towards an embodied conversational agent: effects of the interaction mode. J Pragmatics 2010 Sep;42(9):2385-2397. [doi: 10.1016/j.pragma.2009.12.016]
53. Olafsson S, Kimani E, Asadi R, Bickmore T. That's a rap. In: Intelligent Virtual Agents. Cham: Springer; 2017.
54. Olafsson S, O'Leary T, Bickmore T. Coerced change-talk with conversational agents promotes confidence in behavior change. In: Proceedings of the 13th International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: 13th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy. [doi: 10.1145/3329189.3329202]
55. Parmar D, Olafsson S, Utami D, Bickmore T. Looking the part: the effect of attire and setting on perceptions of a virtual health counselor. In: Proceedings of the 18th ACM International Conference on Intelligent Virtual Agents. 2018 Presented at: 18th ACM International Conference on Intelligent Virtual Agents; Nov 5-8, 2018; Sydney, Australia. [doi: 10.1145/3267851.3267915]
56. Ring L, Shi L, Totzke K, Bickmore T. Social support agents for older adults: longitudinal affective computing in the home. J Multimodal User Interfaces 2014 Jun 18;9(1):79-88. [doi: 10.1007/s12193-014-0157-0]
57. Ring L, Utami D, Bickmore T. The right agent for the job? In: Intelligent Virtual Agents. Cham: Springer; 2014.
58. Ring L, Utami D, Olafsson S, Bickmore T. Increasing engagement with virtual agents using automatic camera motion. In: Intelligent Virtual Agents. Cham: Springer; 2016.
59. Sannon S, Stoll B, DiFranzo D, Jung M, Bazarova N. How personification and interactivity influence stress-related disclosures to conversational agents. In: Proceedings of the Computer Supported Cooperative Work and Social Computing. 2018 Presented at: CSCW '18: Computer Supported Cooperative Work and Social Computing; Nov 3-7, 2018; Jersey City, USA. [doi: 10.1145/3272973.3274076]
60. Shamekhi A, Czerwinski M, Mark G, Novotny M, Bennett G. An exploratory study toward the preferred conversational style for compatible virtual agents. In: Intelligent Virtual Agents. Cham: Springer; 2016.
61. Smestad T, Volden F. Chatbot personalities matters: improving the user experience of chatbot interfaces. In: Internet Science. Cham: Springer; 2019.
62. Straßmann C, Krämer NC, Buschmeier H, Kopp S. Age-related differences in the evaluation of a virtual health agent's appearance and embodiment in a health-related interaction: experimental lab study. J Med Internet Res 2020 Apr 23;22(4):e13726 [FREE Full text] [doi: 10.2196/13726] [Medline: 32324146]
63. van Vugt HC, Konijn E, Hoorn J, Veldhuis J. Why fat interface characters are better e-health advisors. In: Proceedings of the 6th International Conference on Intelligent Virtual Agents, IVA 2006. 2006 Presented at: 6th International Conference on Intelligent Virtual Agents, IVA 2006; Aug 21-23, 2006; CA, United States. [doi: 10.1007/11821830_1]
64. Wissen A, Vinkers C, Halteren A. Developing a virtual coach for chronic patients: a user study on the impact of similarity, familiarity and realism. In: Proceedings of the 11th International Conference on Persuasive Technology - Volume 9638. 2016 Presented at: 11th International Conference on Persuasive Technology; Apr 5 - 7, 2016; Salzburg, Austria. [doi: 10.1007/978-3-319-31510-2_23]
65. Yin L, Bickmore T, Cortes D. The impact of linguistic and cultural congruity on persuasion by conversational agents. In: Intelligent Virtual Agents. Berlin, Heidelberg: Springer; 2010.
66. Zhang Z, Bickmore TW, Paasche-Orlow MK. Perceived organizational affiliation and its effects on patient trust: role modeling with embodied conversational agents. Patient Educ Couns 2017 Sep;100(9):1730-1737. [doi: 10.1016/j.pec.2017.03.017] [Medline: 28381330]
67. Zhou S, Bickmore T, Paasche-Orlow M, Jack B. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In: Intelligent Virtual Agents. Cham: Springer; 2014.
68. Zhou S, Zhang Z, Bickmore T. Adapting a persuasive conversational agent for the Chinese culture. In: Proceedings of the International Conference on Culture and Computing. 2017 Presented at: 2017 International Conference on Culture and Computing; Sep 10-12, 2017; Kyoto, Japan. [doi: 10.1109/culture.and.computing.2017.42]
69. Wang H, Kline JA, Jackson BE, Laureano-Phillips J, Robinson RD, Cowden CD, et al. Association between emergency physician self-reported empathy and patient satisfaction. PLoS One 2018 Sep 13;13(9):e0204113 [FREE Full text] [doi: 10.1371/journal.pone.0204113] [Medline: 30212564]

70. Hojat M, Louis DZ, Maxwell K, Markham FW, Wender RC, Gonnella JS. A brief instrument to measure patients' overall satisfaction with primary care physicians. Fam Med 2011 Jun;43(6):412-417 [FREE Full text] [Medline: 21656396]
71. Derlega V, Winstead B, Greene K. Self-disclosure and starting a close relationship. In: Handbook of relationship initiation. New York, USA: Psychology Press; 2008.
72. Audet CT, Everall RD. Therapist self-disclosure and the therapeutic relationship: a phenomenological study from the client perspective. British J Guidance Counsel 2010 Aug;38(3):327-342. [doi: 10.1080/03069885.2010.482450]
73. Nass C, Steuer J, Tauber E. Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: CHI' 94: ACM Conference on Human Factors in Computing Systems; Apr 24-28, 1994; Boston, MA, USA. [doi: 10.1145/191666.191703]
74. Nass C, Moon Y. Machines and mindlessness: social responses to computers. J Social Isssues 2000 Jan;56(1):81-103. [doi: 10.1111/0022-4537.00153]
75. Mori M, MacDorman K, Kageki N. The uncanny valley [from the field]. IEEE Robot Automat Mag 2012 Jun;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]
76. McTear MF. Spoken dialogue technology. ACM Comput Surv 2002 Mar;34(1):90-169. [doi: 10.1145/505282.505285]
77. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]

Abbreviations

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by R Kukafka; submitted 02.07.21; peer-reviewed by A Mazeas, K Nalin, H Tanaka, R Raeside, V Stara, M Rampioni; comments to author 25.08.21; revised version received 19.09.21; accepted 21.10.21; published 21.12.21.

Please cite as:
Curtis RG, Bartel B, Ferguson T, Blake HT, Northcott C, Virgara R, Maher CA
Improving User Experience of Virtual Health Assistants: Scoping Review
J Med Internet Res 2021;23(12):e31737
URL: https://www.jmir.org/2021/12/e31737
doi:10.2196/31737
PMID:34931997

Review

Digital Interventions to Improve Health Literacy Among Parents of Children Aged 0 to 12 Years With a Health Condition: Systematic Review

Evalotte Mörelius[1,2,3], BSc, MSc, PhD; Suzanne Robinson[1,3], BSc; Diana Arabiat[1,4,5], BSc, MSc, PhD; Lisa Whitehead[1,3], BSc, MSc, PhD

[1]School of Nursing and Midwifery, Edith Cowan University, Joondalup, WA, Australia

[2]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[3]The Centre for Evidence Informed Nursing, Midwifery and Healthcare Practice, Joondalup, WA, Australia

[4]Maternal and Child Nursing Department, Faculty of Nursing, The University of Jordan, Amman, Jordan

[5]Australian Research Council Centre of Excellence for the Digital Child, Joondalup, WA, Australia

Corresponding Author:
Evalotte Mörelius, BSc, MSc, PhD
School of Nursing and Midwifery
Edith Cowan University
270 Joondalup Drive
Joondalup, WA, 6027
Australia
Phone: 61 8 6304 2272
Email: e.morelius@ecu.edu.au

Abstract

Background: Parental health literacy is associated with child health outcomes. Parents are increasingly turning to the internet to obtain health information. In response, health care providers are using digital interventions to communicate information to assist parents in managing their child's health conditions. Despite the emergence of interventions to improve parental health literacy, to date, no systematic evaluation of the effectiveness of the interventions has been undertaken.

Objective: The aim of this review is to examine the effect of digital health interventions on health literacy among parents of children aged 0-12 years with a health condition. This includes evaluating parents' engagement (use and satisfaction) with digital health interventions, the effect of these interventions on parental health knowledge and health behavior, and the subsequent impact on child health outcomes.

Methods: This systematic review was registered a priori on PROSPERO (International Prospective Register of Systematic Reviews) and developed according to the Joanna Briggs Institute methodology for systematic reviews. The databases CINAHL, MEDLINE, and PsycINFO were searched for relevant literature published between January 2010 and April 2021. Studies were included if they were written in English. A total of 2 authors independently assessed the search results and performed a critical appraisal of the studies.

Results: Following the review of 1351 abstracts, 31 (2.29%) studies were selected for full-text review. Of the 31 studies, 6 (19%) studies met the inclusion criteria. Of the 6 studies, 1 (17%) was excluded following the critical appraisal, and the 5 (83%) remaining studies were quantitative in design and included digital health interventions using web-based portals to improve parents' health knowledge and health behavior. Owing to heterogeneity in the reported outcomes, meta-analysis was not possible, and the findings were presented in narrative form. Of the 5 studies, satisfaction was measured in 3 (60%) studies, and all the studies reported high satisfaction with the digital intervention. All the studies reported improvement in parental health literacy at postintervention as either increase in disease-specific knowledge or changes in health behavior. Of the 5 studies, only 1 (20%) study included child health outcomes, and this study reported significant improvements related to increased parental health knowledge.

Conclusions: In response to a pandemic such as COVID-19, there is an increased need for evidence-based digital health interventions for families of children living with health conditions. This review has shown the potential of digital health interventions to improve health knowledge and behavior among parents of young children with a health condition. However, few digital health

interventions have been developed and evaluated for this population. Future studies with robust research designs are needed and should include the potential benefits of increased parent health literacy for the child.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020192386; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=192386

(J Med Internet Res 2021;23(12):e31665) doi:10.2196/31665

KEYWORDS

child; child health services; digital technology; health literacy; infant; internet-based intervention; parents; patient compliance; pediatric hospitals

Introduction

Background

Parents of young children are responsible for the health and well-being of their children and are the advocates and primary caregivers of their children [1]. Parents are expected to interact with health services in the delivery of health care [2] and learn about their child's health condition and potential interventions and procedures involved in treatment [3,4]. To interact effectively with health services, parents of children with health conditions need a level of health literacy. Health literacy increases the parents' capacity to take responsibility and be involved in making decisions related to their child's health [2].

Health literacy is commonly described as knowledge, motivation, and competence to assess, understand, appraise, and apply health information. Health literacy enables people to make decisions on health care, disease prevention, and health promotion throughout their life course [5]. A parent's level of health literacy influences 3 aspects of health behavior: access and use of health services, patient-provider interactions, and self-management [5].

Low parental health literacy has been linked to poor health knowledge and child health status [6-9]. Limited health literacy has been associated with various concerns, including delayed diagnoses, misunderstanding of medication labels [10,11], poor adherence to treatment regimens, and increased use of emergency care [12]. Limited parental health literacy may also result in poor child health outcomes, including depressive symptoms, persistent asthma [8,13-16], and less than optimal glycemic control in children with diabetes [17].

Parents' need for health information can remain unmet [18]. Moreover, parents of children with health conditions are often affected by stress [19] and poor sleep, which may hinder their ability to receive and process new information and learn about their child's condition to provide them with optimal care [20,21]. Information about a child's health condition needs to be delivered at a time that is appropriate for the family and offered as many times as needed, which can be challenging for both the parents and the health service [22,23]. A way to enhance the delivery of information is through digital technology (eg, mobile phones and tablets), as this can be accessed at a time, place, and pace that best suits the parent. Thus, digital health interventions (eg, information videos, web-based platforms, and mobile apps) can be used in the home settings where most care is provided [3]. Internet and mobile phone use are high among parents worldwide, with most parents using the internet and mobile phones to access information multiple times a day [24,25]. Parents are heavy users of web-based child health-related information [25]. Between 70% and 80% of parents have searched on the web for health information, with most parents seeking parenting advice, health information, or social support [26].

Health care providers increasingly use digital technologies to communicate information to address health needs and deliver health care interventions [27]. Despite minimal evidence for the effectiveness of digital health interventions, these have significantly increased because of the restrictions related to COVID-19 [28].

Considerable challenges exist related to how digital health technologies can best be used and integrated to support and facilitate user engagement and provide individualized health information and care [29]. A level of engagement is required to affect a change in health behaviors [30]. Engagement is measured by the extent of use of the intervention (initial log-in and number of activities completed) and the subjective user experience, which is often measured by user satisfaction [31,32].

Objectives

Despite the growth in the use of the internet among parents to obtain health information and the emergence of digital health interventions to support parents, systematic evaluation of parental engagement with and effectiveness of digital interventions to improve health literacy has not been conducted. The objective of this review is to examine the effect of digital health interventions on health literacy among parents of children aged 0-12 years with a health condition. In this study, health literacy includes the evaluation of parental engagement (use and satisfaction) with digital health interventions, the effect of these interventions on parental health knowledge and health behavior, and the subsequent impact on child health outcomes.

Methods

This review followed the Joanna Briggs Institute (JBI) methodology for systematic reviews [33] and was conducted according to the registered a priori PROSPERO (International Prospective Register of Systematic Reviews) protocol CRD42020192386. Qualitative and quantitative studies were included as a broad search was required to address the complex health system–related questions [34].

Inclusion Criteria

Participants

This review considered studies that included parents or primary caregivers of children aged 0-12 years with a health condition. Health conditions could be an acute or a chronic disease, diagnosis, or condition. Studies involving children aged ≥12 years were excluded, as children aged ≥12 years are more likely and encouraged to take more responsibility for their health care. Studies were also excluded if health-related information was directed to the child, if the child was healthy, or if the primary users of the digital health intervention were health professionals such as medical staff, nursing staff, health care management, and administrators or researchers. Studies focusing on health promotion or disease prevention (eg, increased knowledge of obesity, physical activity, and smoking) were excluded.

Intervention

This review considered studies that examined any digital health intervention that aimed to improve health literacy among parents. Interventions could focus on communication (eg, web-based platforms, mobile apps, videoconferencing, and SMS text messaging), education (eg, videos, web-based platforms, mobile apps, and interactive training), or a combination of communication and education interventions. Interventions that targeted the child or clinician or were delivered directly (eg, face to face by health care professionals) to the parent were excluded.

Context

This review considered studies that examined the use of digital health interventions in both the home and hospital setting.

Outcomes

This review considered studies that included an increase or decrease in health literacy defined by health knowledge or health behavior [5,30]. The review also considered the following outcomes: changes in the child's health outcome and engagement with the digital health intervention, including (1) use, that is, *logging into* the web-based platform, *continued use of* the digital health intervention, amount and duration of access, and type of information accessed; or (2) satisfaction with the digital health intervention measured, for example, by attention, interest, usefulness, and perceived benefits through survey items [31,32].

Search Strategy

The following databases were searched for existing systematic reviews on this topic: CINAHL, MEDLINE, PROSPERO, JBI Database of Systematic Reviews and Implementation Reports, and Cochrane Library Database of Systematic Reviews, and no reviews were located on the specific topic. The base search strategy was developed in CINAHL, and additional adapted searches were run on CINAHL, MEDLINE, and PsycINFO and hand searched in Google Scholar. The search strategy was developed in collaboration with a research librarian to identify articles examining the health literacy of parents of children with a health condition. The key terms were *health literacy* or *health behavior* or *health education* or *health information* and *digital health* or *mobile health* or *electronic health*. The initial search was undertaken in June 2020 and updated in April 2021. The search was limited to papers published in English between January 2010 and April 2021, as there have been constant developments and improvements in digital health technology over the past 10 years. The complete search strategy is shown in Multimedia Appendix 1. The reference list of all the included studies was reviewed to identify any relevant papers not found in the electronic search. Gray literature sources including OpenGrey, ProQuest Dissertation and Theses, Google, and Google Scholar were also searched to identify unpublished studies.

Screening of Articles

After removing duplicates using Endnote (Clarivate), the web-based tool Rayyan (Rayyan Systems Inc) was used to screen the articles [35]. Titles and abstracts were reviewed independently by 2 authors (EM and SR) to determine if they met the inclusion criteria. Any disagreement regarding eligibility was resolved through discussion with a third author (DA). The selected articles were reviewed in full text by the same authors.

Assessment of Methodological Quality

The quality of the screened studies was critically appraised independently by 2 reviewers (EM and SR) using the Mixed Methods Appraisal Tool (MMAT) version 2018 [36]. The MMAT was selected in preference to the JBI critical appraisal tool as the MMAT was developed to appraise studies that combine qualitative, quantitative, and mixed methods studies.

Data Extraction

Data extraction was undertaken by the second author (SR) and reviewed by the first author (EM) and modeled on the standardized data extraction tool from the JBI [33]. The extracted data included specific details about the study setting and context; phenomena of interest (health literacy); study design; sampling of participants, sample size, and characteristics of the study sample; specific details about the interventions; and outcomes of significance to the review question. All data were extracted following a thorough, complete reading of the text to identify qualitative and quantitative findings relevant to the objectives and questions of the review.

Data Synthesis

Owing to differences in reported quantitative data and the small number of studies, meta-analysis was not possible. The findings are presented in a narrative form [37], including tables and figures to aid in data presentation.

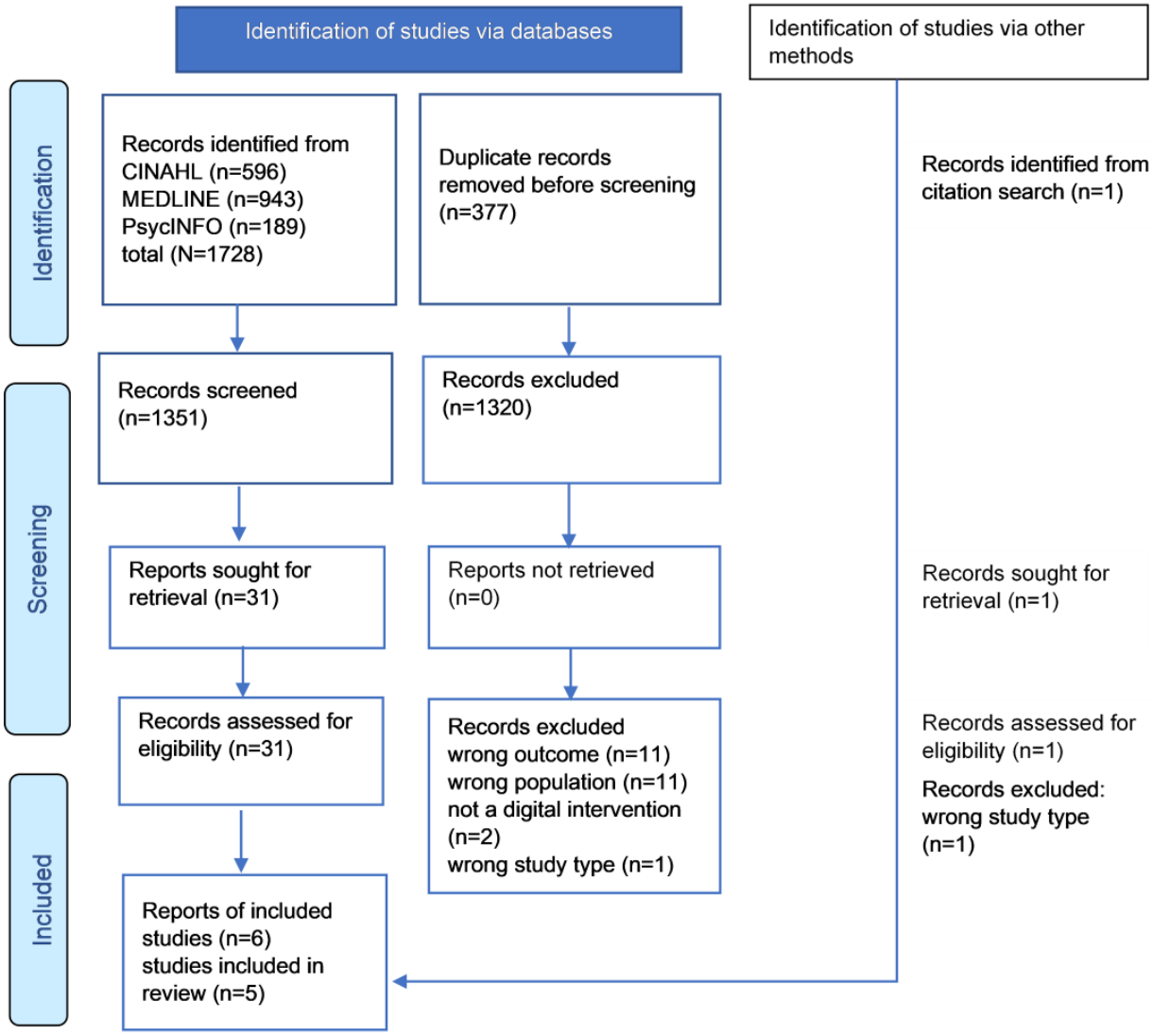
Results

Study Inclusion

In total, 1728 references were identified using the search terms. The addition of secondary searches of reference lists and searches of gray literature resulted in the identification of another reference. The exclusion of duplicates resulted in 1351 references, of which 1320 (97.71%) were excluded after the title and abstract screening. The remaining 31 references were retrieved in full text. Of the 31 papers, 25 (81%) were excluded

(Multimedia Appendix 2 [38-62]), resulting in 6 (19%) papers eligible for inclusion (Figure 1) [63-68].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study selection and inclusion process.



Methodological Quality

A total of 2 authors (EM and SR) independently appraised the 6 articles that met the inclusion criteria for methodological quality. Table 1 summarizes the questions answered in the MMAT, with 1 study excluded as it did not meet the essential screening criteria questions [36,68]. The methods varied across the remaining 5 studies. Of the 5 studies, 2 (40%) studies used mixed methods [63,64] and 3 (60%) studies used a quasi-experimental design (quantitative nonrandomized) [65-67].

Table 1. Assessment of methodological quality with the Mixed Methods Appraisal Tool version 2018 [36].

	Blatz et al [63]	Fiks et al [64]	Kobak et al [65]	McGarry et al [66]	Ruiz-Baqués et al [67]	Slater et al [68]
Screening questions						
Are there clear research questions?	Yes	Yes	Yes	Yes	Yes	Unclear
Do the collected data allow to address the research questions?	Yes	Yes	Yes	Yes	Yes	Unclear
Quantitative descriptive						
Is the sampling strategy relevant to address the research question?	Yes	Yes	N/A[a]	N/A	N/A	N/A
Is the sample representative of the target population?	Yes	Yes	N/A	N/A	N/A	N/A
Are the measurements appropriate?	No	Yes	N/A	N/A	N/A	N/A
Is the risk of nonresponse bias low?	Yes	Yes	N/A	N/A	N/A	N/A
Is the statistical analysis appropriate to answer the research question?	Yes	Yes	N/A	N/A	N/A	N/A
Quantitative nonrandomized						
Are the participants representative of the target population?	N/A	N/A	No	No	Yes	N/A
Are measurements appropriate regarding both the outcome and intervention (or exposure)?	N/A	N/A	Yes	Yes	No	N/A
Are there complete outcome data?	N/A	N/A	Yes	No	Yes	N/A
Are the confounders accounted for in the design and analysis?	N/A	N/A	Yes	Yes	Yes	N/A
During the study period, was the intervention administered (or exposure occurred) as intended?	N/A	N/A	Yes	Yes	Yes	N/A
Quantitative score, n (%)	4 (80)	5 (100)	4 (80)	3 (60)	4 (80)	N/A
Qualitative						
Is the qualitative approach appropriate to answer the research question?	Yes	Yes	N/A	N/A	N/A	N/A
Are the qualitative data collection methods adequate to address the research question?	No	No	N/A	N/A	N/A	N/A
Are the findings adequately derived from the data?	No	No	N/A	N/A	N/A	N/A
Is the interpretation of results sufficiently substantiated by the data?	No	No	N/A	N/A	N/A	N/A
Is there coherence between qualitative data sources, collection, analysis, and interpretation?	No	No	N/A	N/A	N/A	N/A
Qualitative score, n (%)	1 (20)[b]	1 (20)[b]	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

[b]Only the quantitative part was used in the review as the qualitative information provided was minimal.

The quality of the 60% (3/5) quasi-experimental studies was moderate, with a 60% to 80% score [65-67], as was the quality of the quantitative component of the mixed methods studies, which was moderate to good with an 80% to 100% score [63,64]. The quality of the qualitative component of the 40% (2/5) mixed methods studies was poor (20%) and was therefore excluded [63,64].

Although the 5 included studies met the quality criteria, biases were noted. A mixed methods study [63] and a quasi-experimental study [67] used nonvalidated measurement tools. In 67% (2/3) of the quasi-experimental studies, it was unclear if the participants were representative of the population, as a large proportion of participants were White [65] or college-educated [66]. McGarry et al [66] did not provide complete outcome data, with a high rate of participants not

completing the program and not providing a rationale for why they dropped out of the program [66]. None of the 5 studies were randomized or had an independent control group, limiting the ability to quantify the effect of the digital health intervention and limit the confounding factors. Of the 5 studies, 3 (60%) had small sample sizes, with not more than 30 participants in each study receiving the digital health intervention [63,65,66]. Of these 3 studies, 2 (67%) were pilot studies [65,66].

Characteristics of the Studies

Of the 5 studies, 2 (40%) used a descriptive longitudinal design with repeated measures [63,64] and 3 (60%) used a pretest–posttest design [65-67]. Of the 5 studies identified as meeting the selection criteria, 1 (20%) study was implemented in a hospital [63], 2 (40%) in outpatient clinics [64,65], and 2 (40%) in the community [66,67]. Of the 5 studies, 4 (80%) were conducted in the United States [63-66] and 1 (20%) study in Spain [67]. More detailed information about the study characteristics is provided in Multimedia Appendix 3 [63-67].

Description of the Participants

The sample size varied across the 5 studies, with 3 (60%) studies having <30 participants [63,65,66] and 2 (40%) studies with >200 participants [64,67]. Most participants were mothers, ranging from 73% [66] to 100% of the sample in each study [63]. The mean age of participants in each study varied; the lowest mean age was 28.6 years [63], and the highest was 37.5 years [64]. A higher percentage of male children living with health conditions was represented in the studies, ranging from 54.5% [64] to 73% [66] (Multimedia Appendix 3).

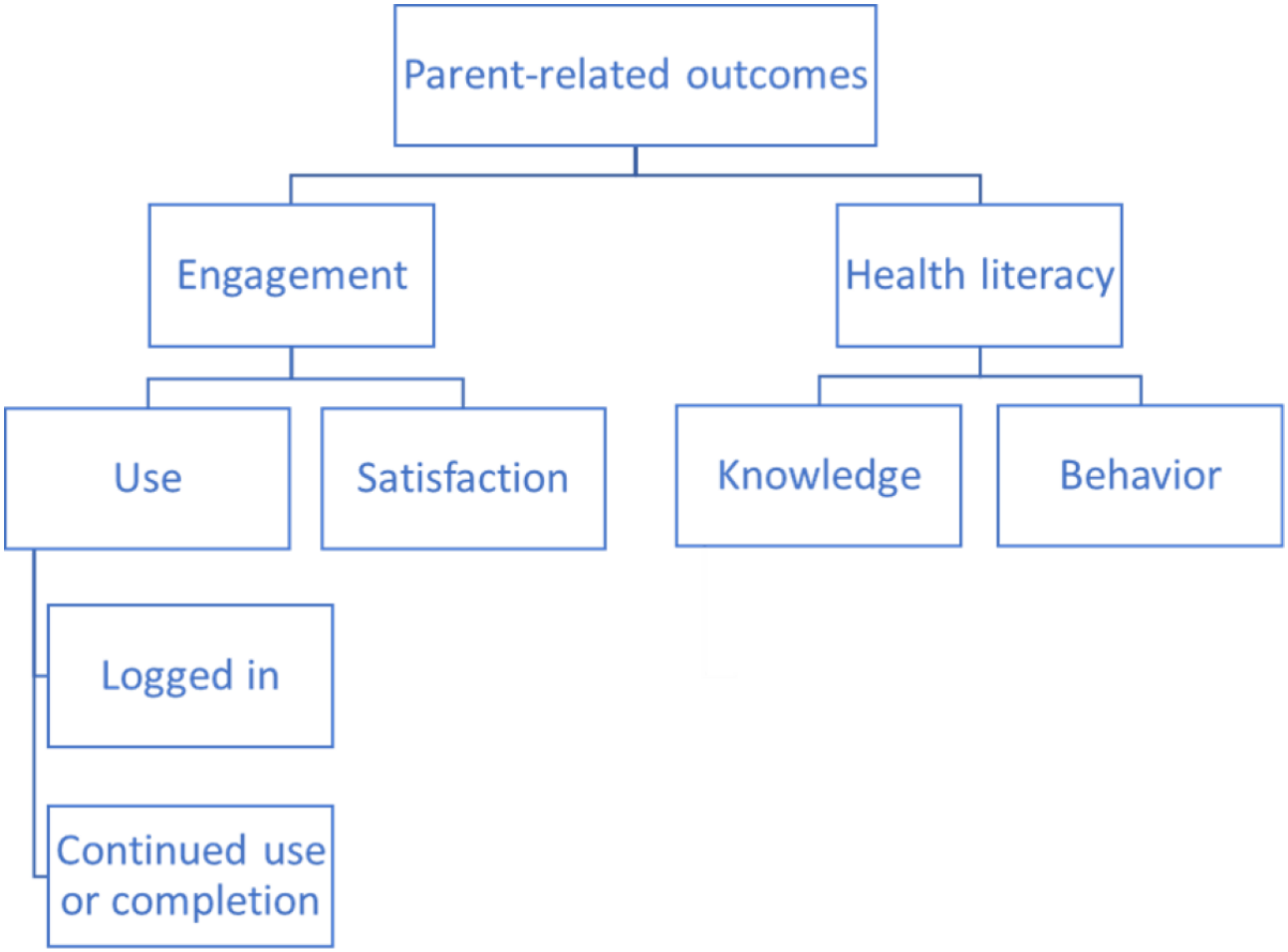
Description of the Interventions

All the 5 digital interventions used web-based portals, with access limited to the participants only (ie, access was not open to the general public). The interventions targeted a variety of health conditions: infants born preterm [63], asthma [64], autism spectrum disorder [65,66], and food allergy [67]. All 5 study interventions included an educational component, with 4 (80%) studies including additional interactive communication components (electronic recording of times and volumes of breastmilk expression, a patient portal to interact with health care providers [64], use of videos, web-based training of parents to promote the child's communication skills [66], and web-based discussion forums [67]; Multimedia Appendix 4 [63-67]).

Parent-Related Outcomes

Parent-related outcomes were extracted as engagement (use and satisfaction) and health literacy (health knowledge and health behavior; Figure 2).

Figure 2. Parent-related outcomes extracted as engagement and health literacy, including subcontent.



Engagement

All studies reported on engagement, although the definitions of engagement and levels of engagement varied across the studies. Additional details are outlined in Multimedia Appendix 4 and further described in the following sections.

Use

Approximately 80% (4/5) of the studies reported the number of parents who were invited to participate [63,64,66,67]. Invitees logging into the website ranged from 2.6% of eligible parents [64] to 100% [66] (Table 2). Most participants (67%-100%) who logged into the website accessed the digital intervention at least once [63-67]. Users who continued to use the intervention or completed the digital health intervention ranged from 37% of parents participating in the Pivotal Response Treatment (PRT) program [66] to 100% in the Lactation Log Plus website [63] and the web-based tutorial, which focused on communication with children with autism spectrum disorder [65] (Table 2).

Table 2. Number and percentage of parents approached to participate, logged into the website, and accessed the digital intervention (use).

Study	Approached to participate, N	Logged into the website, n (%)	Accessed the intervention, n (%)	Continued use or completed the intervention, n (%)
Blatz et al [63]	20	18 (90)	13 (100)	13 (100)
Fiks et al [64]	9133	237 (2.59)	237 (100)	156 (65.8)
Kobak et al [65]	—[a]	23 (—)	23 (100)	23 (100)
McGarry et al [66]	51	30 (59)	30 (100)	11 (37)
Ruiz-Baqués et al [67]	277	207 (74.5)	139 (67.1)	130 (62.8)

[a]Not provided.

Satisfaction

Of the 5 studies, 3 (60%) measured satisfaction with digital health interventions. Tools used to measure satisfaction varied, with high satisfaction identified in all 3 studies [65-67].

Kobak et al [65] used the System Usability Scale to measure satisfaction with the technical parts of the web-based version of the Enhancing Interaction Program. This is a validated 10-item scale ranging from 0 to 100. The mean score in this study was 85 (SD 17), which corresponds to a score of excellent. They also used the User Satisfaction Questionnaire to evaluate the clinical content of the web-based program. This scale ranges from 15 to 60 and has shown good internal consistency. The mean score in this study was 54.5 (SD 5.9). McGarry et al [66] used Social Validity Measures to assess parents' satisfaction with the PRT program. Parents were asked to respond to a variety of statements using a scale from 0 (strongly disagree) to 5 (strongly agree). Over 90% of the parents agreed that the course was well-written and organized. Ruiz-Baqués et al [67] used a 5-item Likert scale to assess satisfaction with the educational program. The scale ranged from 0 (not at all) to 10 (great deal) points. The mean score in this study was 8.78.

Health Literacy

All studies reported on health literacy, either as a change in health knowledge or health behavior. Additional details are outlined in Multimedia Appendix 4 and further described in the following sections.

Health Knowledge

Improvement in parental knowledge was identified in 40% (2/5) of the studies [65,67]. Kobak et al [65] used a questionnaire to measure changes in knowledge before and after a web-based intervention program for autism spectrum disorder and found an increase in the mean number of correct items from 12.6 to 20.4 (P<.001); 79% of parents scored ≥80% after taking the tutorial compared with 8% before taking the tutorial. Ruiz-Baqués et al [67] measured the changes in knowledge before and after a web-based intervention program for food allergy and found an improvement in 15 out of 30 questionnaire items and a significant improvement (P<.001) in 8 items. Improvement was more frequent in the *general knowledge and clinical aspects* domain than in the *daily life with food allergy* domain. No study reported on the time elapsed between the end of the intervention and the knowledge test.

Health Behavior

Of the 5 studies, improvement in behavior was identified in 3 (60%) studies [63,64,66]. Blatz et al [63] found that a website that included breast milk information and a milk diary helped participants pump milk and sustain milk supply. Of the 13 participants, 2 (15%) felt that the milk log website helped to pump milk a great deal, 5 (38%) felt that it somewhat helped, and 6 (46%) felt that the website log did not help them pump breast milk. Furthermore, of the 13 participants, 2 (15%) felt that the milk log website helped maintain milk supply a great deal, 3 (23%) felt that it somewhat helped, and 8 (62%) felt that the website log did not help them maintain it.

Fiks et al [64] found increased medication refills and asthma-related medical visits among parents of children with uncontrolled asthma. Of the 76 children with uncontrolled asthma after the first survey, 20 (67%) had a medication change or refill within 30 days of survey completion, and 21 (28%) had an asthma-related primary care visit within 30 days. The results represent a significant increase in medication changes or refills and asthma-related visits when compared with the same period in the prior year for each child (14% increase in medication changes, 95% CI 2%-27%, and 16% increase in visits, 95% CI 3%-28%).

McGarry et al [66] found that parents who completed the Autism PRT program were successfully able to learn and implement

the strategies. Parents submitted videos of parent-child interactions, which were coded for the fidelity of implementation and social communication behaviors. Parent's treatment fidelity improved from baseline (mean 65.34%, SD 18.04%) to week 5 (mean 90.13%, SD 7.20%; *P*<.001). At baseline, of the 11 parents, 1 (9%) met the fidelity of implementation, with ≥80% fidelity score. By the end of the program, 10 parents met the fidelity of implementation, whereas 1 parent approached fidelity with a 75% fidelity score.

Child Health Outcomes

Of the 5 studies, 1 study (20%) reported on child health outcomes. McGarry et al [66] demonstrated increased communication behavior among children with autism spectrum disorder. There was an improvement from baseline to week 5 in children's vocalization (*P*=.05), eye contact (*P*=.03), and positive affect (*P*<.001).

Discussion

Principal Findings

With only 5 intervention studies identified in this systematic review, it is clear that few digital health interventions have been developed to improve health literacy among parents of children aged 0 to 12 years living with a health condition. Of the 5 studies, 4 (80%) were published in the past 4 years, suggesting that digital health interventions to improve health literacy are an emerging area of research. The use of digital health interventions in clinical practice has also increased because of the COVID-19 pandemic [25].

The 5 studies reported parent-related outcomes, with a focus on engagement with digital health interventions. Despite the low number of studies, digital health interventions to improve health literacy appear acceptable and useful among parents. After parents logged into a website and enrolled in the digital health intervention, >60% of parents demonstrated continued engagement with the program [63-65,67], with 1 exception [66]. The study by McGarry et al [66] had the lowest rate of continued use. The reason could be a rather demanding program where the parents had to submit a video after each web-based lesson, capturing how they used their new information with the child. Submission of a video was a prerequisite for continuing with the intervention. In another study, <3% of the parents who were approached logged into the intervention platform [64]. However, in the same study, approximately 66% of the parents continued to use the portal once logged in [64], implying that parents who initiate engagement are also likely to remain engaged. A systematic review studying the factors affecting engagement and recruitment to digital health interventions suggests that people struggle to make sense of digital health interventions and recommends raising the profile for digital health products to make people more aware of them [69]. It is possible that the current pandemic, which has increased the use of digital health care, has raised the awareness of digital interventions.

Improvement in parental health literacy, including either knowledge [65,67] or behavior change [63,64,66], indicated positive results, which are important when parents are responsible for their children's health care and well-being [1]. However, none of the included studies used a validated instrument developed to measure general health literacy, such as Rapid Estimate of Adult Literacy in Medicine or Test of Functional Health Literacy in Adults [70]. Instead, both Kobak et al [65] and Ruiz-Baqués et al [67] measured specific knowledge targeting autism spectrum disorder and food allergy, respectively. Disease-specific knowledge has proved to be of greater importance in affecting health behavior change than general health literacy [71]. It is also suggested that general health literacy can be a prerequisite for disease-specific knowledge [71]. Health literacy is a complex concept in which knowledge is important and can influence health behavior and child health outcomes [5]. For parents responsible for their children's basic care and specific care related to their health condition, it is essential to have both general health literacy and disease-specific knowledge. However, few studies have evaluated the effect of digital health interventions on either parents' general health literacy or disease-specific knowledge. In times of a pandemic, when access to physical consultations with health care providers has been affected, digital health interventions to increase parents' knowledge and behavior may be of utmost importance for the child's health.

Health literacy increases the parent's capacity to engage in and take responsibility for their child's health care [2], both of which are associated with improved child health outcomes. Only 20% (1/5) of the studies reported on a change in outcomes [66]. McGarry et al [66] reported positive changes in children's communicative behavior at the same time as the parent's treatment fidelity improved significantly. These are promising results showing that increased disease-specific knowledge in parents can positively affect child health outcomes; however, more studies are necessary to prove this hypothesis. Therefore, it is essential to report on potential improvements in the child's health status in future studies to evaluate the impact of increased health literacy in parents.

No randomized controlled studies were included in this review. Of the 5 studies, 2 (40%) were presented as mixed methods studies [63,64]. Unfortunately, the qualitative part of these studies was assessed as having low methodological quality. The only qualitative study included following full-text review was subsequently excluded after critical appraisal [68]. The 3 quantitative studies that evaluated satisfaction reported high satisfaction levels with the digital health intervention [65-67]. However, these studies all used different tools to measure satisfaction, thus limiting the aggregation and synthesis of data. Qualitative studies could help to develop an understanding of why some parents decide to initiate engagement with digital health interventions and what factors contribute to their continued use.

Limitations of the Review

Although this review was systematic, the findings must be interpreted with caution. The number of studies identified was small, homogeneity among the studies was limited, and none of the included studies used a true comparison (ie, control group). Several factors may have influenced the outcome of the digital health interventions, including the implementation methodology, limited responses and participation rates,

encouragement by health care providers, and participants' characteristics. Owing to the small number of included studies and missing data on participants' characteristics, the influence of potential covariates could not be further evaluated. The variation in methodological design, including differences in outcomes and few comparators, limited the authors' ability to conduct a meta-analysis. It should also be considered that only studies published in English were included in this review. Despite these limitations, the favorable results across studies suggest that further evaluations of the benefits of digital health interventions should be undertaken.

Conclusions

This review has shown the potential of digital health interventions to improve health knowledge and health behavior among parents of children aged 0-12 years with a health condition. Of the 5 included studies, 4 (80%) were published in the past 4 years, indicating that digital health interventions aimed at improving health literacy are a developing research area. Future studies should include qualitative studies and studies with randomized samples to more fully understand the potential of digital health interventions to increase parent health literacy.

Acknowledgments

The authors acknowledge the support from the Australian Research Council Centre of Excellence for the Digital Child.

Authors' Contributions

EM, SR, DA, and LW were involved in conceptualization; EM and SR formulated the methodology and performed the formal analysis; SR wrote the original draft; EM, DA, and LW were involved in the review and editing process; and EM, SR, DA, and LW granted the final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - jmir_v23i12e31665_app1.docx]

Multimedia Appendix 2
Excluded studies.
[DOCX File , 35 KB - jmir_v23i12e31665_app2.docx]

Multimedia Appendix 3
Descriptive information.
[DOCX File , 24 KB - jmir_v23i12e31665_app3.docx]

Multimedia Appendix 4
Description of interventions and outcomes.
[DOCX File , 26 KB - jmir_v23i12e31665_app4.docx]

References

1. Boshoff K, Gibbs D, Phillips RL, Wiles L, Porter L. Parents' voices: 'why and how we advocate'. A meta-synthesis of parents' experiences of advocating for their child with autism spectrum disorder. Child Care Health Dev 2016 Nov;42(6):784-797. [doi: 10.1111/cch.12383] [Medline: 27445227]
2. Aarthun A, Øymar KA, Akerjordet K. How health professionals facilitate parents' involvement in decision-making at the hospital: a parental perspective. J Child Health Care 2018 Mar;22(1):108-121. [doi: 10.1177/1367493517744279] [Medline: 29212362]
3. Smith J, Swallow V, Coyne I. Involving parents in managing their child's long-term condition-a concept synthesis of family-centered care and partnership-in-care. J Pediatr Nurs 2015;30(1):143-159. [doi: 10.1016/j.pedn.2014.10.014] [Medline: 25458112]
4. Smith J, Kendal S. Parents' and health professionals' views of collaboration in the management of childhood long-term conditions. J Pediatr Nurs 2018;43:36-44. [doi: 10.1016/j.pedn.2018.08.011] [Medline: 30473155]
5. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]

6. de Buhr E, Tannen A. Parental health literacy and health knowledge, behaviours and outcomes in children: a cross-sectional survey. BMC Public Health 2020 Jul 13;20(1):1096 [FREE Full text] [doi: 10.1186/s12889-020-08881-5] [Medline: 32660459]
7. Berkman N, Sheridan S, Donahue K, Halpern D, Viera A, Crotty K, et al. Health literacy interventions and outcomes: an updated systematic review. Evid Rep Technol Assess (Full Rep) 2011 Mar(199):1-941. [Medline: 23126607]
8. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics 2009 Nov;124 Suppl 3:S265-S274. [doi: 10.1542/peds.2009-1162B] [Medline: 19861480]
9. Sanders LM, Federico S, Klass P, Abrams MA, Dreyer B. Literacy and child health: a systematic review. Arch Pediatr Adolesc Med 2009 Feb;163(2):131-140. [doi: 10.1001/archpediatrics.2008.539] [Medline: 19188645]
10. Lenahan JL, McCarthy DM, Davis TC, Curtis LM, Serper M, Wolf MS. A drug by any other name: patients' ability to identify medication regimens and its association with adherence and health outcomes. J Health Commun 2013;18 Suppl 1(Suppl 1):31-39 [FREE Full text] [doi: 10.1080/10810730.2013.825671] [Medline: 24093343]
11. Yin HS, Mendelsohn AL, Wolf MS, Parker RM, Fierman A, van Schaick L, et al. Parents' medication administration errors: role of dosing instruments and health literacy. Arch Pediatr Adolesc Med 2010 Feb;164(2):181-186. [doi: 10.1001/archpediatrics.2009.269] [Medline: 20124148]
12. Yin HS, Dreyer BP, Foltin G, van Schaick L, Mendelsohn AL. Association of low caregiver health literacy with reported use of nonstandardized dosing instruments and lack of knowledge of weight-based dosing. Ambul Pediatr 2007;7(4):292-298. [doi: 10.1016/j.ambp.2007.04.004] [Medline: 17660100]
13. Harrington KF, Zhang B, Magruder T, Bailey WC, Gerald LB. The impact of parent's health literacy on pediatric asthma outcomes. Pediatr Allergy Immunol Pulmonol 2015 Mar 01;28(1):20-26 [FREE Full text] [doi: 10.1089/ped.2014.0379] [Medline: 25852967]
14. Shone LP, Conn KM, Sanders L, Halterman JS. The role of parent health literacy among urban children with persistent asthma. Patient Educ Couns 2009 Jun;75(3):368-375 [FREE Full text] [doi: 10.1016/j.pec.2009.01.004] [Medline: 19233588]
15. Morrison AK, Glick A, Yin HS. Health literacy: implications for child health. Pediatr Rev 2019 Jun;40(6):263-277. [doi: 10.1542/pir.2018-0027] [Medline: 31152099]
16. Tzeng Y, Chiang B, Chen Y, Gau B. Health literacy in children with asthma: a systematic review. Pediatr Neonatol 2018 Oct;59(5):429-438 [FREE Full text] [doi: 10.1016/j.pedneo.2017.12.001] [Medline: 29678410]
17. Pulgarón ER, Sanders LM, Patiño-Fernandez AM, Wile D, Sanchez J, Rothman RL, et al. Glycemic control in young children with diabetes: the role of parental health literacy. Patient Educ Couns 2014 Jan;94(1):67-70 [FREE Full text] [doi: 10.1016/j.pec.2013.09.002] [Medline: 24091252]
18. Arabiat D, Whitehead L, Foster M, Shields L, Harris L. Parents' experiences of Family Centred Care practices. J Pediatr Nurs 2018;42:39-44. [doi: 10.1016/j.pedn.2018.06.012] [Medline: 30219298]
19. Cousino MK, Hazen RA. Parenting stress among caregivers of children with chronic illness: a systematic review. J Pediatr Psychol 2013 Sep;38(8):809-828. [doi: 10.1093/jpepsy/jst049] [Medline: 23843630]
20. Angelhoff C, Edéll-Gustafsson U, Mörelius E. Sleep of parents living with a child receiving hospital-based home care: a phenomenographical study. Nurs Res 2015;64(5):372-380. [doi: 10.1097/NNR.0000000000000108] [Medline: 26325279]
21. Mörelius E, Hemmingsson H. Parents of children with physical disabilities - perceived health in parents related to the child's sleep problems and need for attention at night. Child Care Health Dev 2014 May;40(3):412-418. [doi: 10.1111/cch.12079] [Medline: 23731355]
22. Thompson D, Leach M, Smith C, Fereday J, May E. How nurses and other health professionals use learning principles in parent education practice: a scoping review of the literature. Heliyon 2020 Mar 18;6(3):e03564 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03564] [Medline: 32211543]
23. Kepreotes E, Keatinge D, Stone T. The experience of parenting children with chronic health conditions: a new reality. J Nurs Healthc Chronic Illn 2010 Feb 9;2(1):51-62. [doi: 10.1111/j.1752-9824.2010.01047.x]
24. Meyers N, Glick AF, Mendelsohn AL, Parker RM, Sanders LM, Wolf MS, et al. Parents' use of technologies for health management: a health literacy perspective. Acad Pediatr 2020;20(1):23-30 [FREE Full text] [doi: 10.1016/j.acap.2019.01.008] [Medline: 30862511]
25. Kubb C, Foran HM. Online health information seeking by parents for their children: systematic review and agenda for further research. J Med Internet Res 2020 Aug 25;22(8):e19985 [FREE Full text] [doi: 10.2196/19985] [Medline: 32840484]
26. Jaks R, Baumann I, Juvalta S, Dratva J. Parental digital health information seeking behavior in Switzerland: a cross-sectional study. BMC Public Health 2019 Feb 21;19(1):225 [FREE Full text] [doi: 10.1186/s12889-019-6524-8] [Medline: 30791927]
27. Shorey S, Chee CY, Ng ED, Lau Y, Dennis C, Chan YH. Evaluation of a technology-based peer-support intervention program for preventing postnatal depression (part 1): randomized controlled trial. J Med Internet Res 2019 Aug 29;21(8):e12410 [FREE Full text] [doi: 10.2196/12410] [Medline: 31469084]
28. 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]
29. McCabe C, Timmins F. Embracing healthcare technology - What is the way forward for nurse education? Nurse Educ Pract 2016 Nov;21:104-106. [doi: 10.1016/j.nepr.2016.10.007] [Medline: 27810650]

30. 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]
31. 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;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
32. 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]
33. JBI manual for evidence synthesis. JBI Global. 2020. URL: https://doi.org/10.46658/JBIMES-20-01 [accessed 2021-12-04]
34. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Glob Health 2019 Jan 25;4(Suppl 1):e000893 [FREE Full text] [doi: 10.1136/bmjgh-2018-000893] [Medline: 30775016]
35. 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]
36. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
37. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. JBI Manual for Evidence Synthesis. URL: https://jbi-global-wiki.refined.site/space/MANUAL/3283910685/Chapter+3%3A+Systematic+reviews+of+effectiveness [accessed 2021-12-04]
38. Armstrong-Heimsoth A, Johnson M, McCulley A, Basinger M, Maki K, Davison D. Good googling: a consumer health literacy program empowering parents to find quality health information online. J Consum Health Internet 2017 Jun 12;21(2):111-124. [doi: 10.1080/15398285.2017.1308191]
39. Ayre J, Costa DS, McCaffery KJ, Nutbeam D, Muscat DM. Validation of an Australian parenting health literacy skills instrument: the parenting plus skills index. Patient Educ Couns 2020 Jun;103(6):1245-1251. [doi: 10.1016/j.pec.2020.01.012] [Medline: 31982204]
40. Asan O, Scanlon MC, Crotty B, Holden RJ, Flynn KE. Parental perceptions of displayed patient data in a PICU: an example of unintentional empowerment. Pediatr Crit Care Med 2019 May;20(5):435-441 [FREE Full text] [doi: 10.1097/PCC.0000000000001895] [Medline: 31058783]
41. Byczkowski TL, Munafo JK, Britto MT. Family perceptions of the usability and value of chronic disease web-based patient portals. Health Informatics J 2014 Jun;20(2):151-162 [FREE Full text] [doi: 10.1177/1460458213489054] [Medline: 24056751]
42. Chau S, Oldman S, Smith SR, Lin CA, Ali S, Duffy VB. Online behavioral screener with tailored obesity prevention messages: application to a pediatric clinical setting. Nutrients 2021 Jan 14;13(1):223 [FREE Full text] [doi: 10.3390/nu13010223] [Medline: 33466705]
43. Chorianopoulou A, Lialiou P, Mechili EA, Mantas J, Diomidous M. Investigation of the quality and effectiveness of telemedicine in children with diabetes. Stud Health Technol Inform 2015;210:105-109. [Medline: 25991111]
44. Dudovitz R, Teutsch C, Holt K, Herman A. Improving parent oral health literacy in Head Start programs. J Public Health Dent 2020 Jun;80(2):150-158. [doi: 10.1111/jphd.12361] [Medline: 32083725]
45. Edwards C, Bolling-Walker K, Deupree J. Actionability and usability of a fever management tool for pediatric caregivers. J Contin Educ Nurs 2020 Jul 01;51(7):338-344. [doi: 10.3928/00220124-20200611-10] [Medline: 32579230]
46. Fagnano M, Halterman JS, Conn KM, Shone LP. Health literacy and sources of health information for caregivers of urban children with asthma. Clin Pediatr (Phila) 2012 Mar;51(3):267-273 [FREE Full text] [doi: 10.1177/0009922811421413] [Medline: 21911409]
47. Fauer AJ, Hoodin F, Lalonde L, Errickson J, Runaas L, Churay T, et al. Impact of a health information technology tool addressing information needs of caregivers of adult and pediatric hematopoietic stem cell transplantation patients. Support Care Cancer 2019 Jun;27(6):2103-2112 [FREE Full text] [doi: 10.1007/s00520-018-4450-4] [Medline: 30232587]
48. Gage-Bouchard EA, LaValley S, Mollica M, Beaupin LK. Communication and exchange of specialized health-related support among people with experiential similarity on facebook. Health Commun 2017 Oct;32(10):1233-1240. [doi: 10.1080/10410236.2016.1196518] [Medline: 27485860]
49. Guðmundsdóttir K, Ala'i-Rosales S, Sigurðardóttir Z. Extending caregiver training via telecommunication for rural Icelandic children with autism. Rural Special Educ Q 2018 Jul 19;38(1):26-42. [doi: 10.1177/8756870518783522]
50. Kaskinen A, Ayeboa-Sallah B, Teivaanmäki T, Wärnhjelm E, Korhonen L, Helve O. Pediatric web-based chat services for caregivers of children: descriptive study. J Med Internet Res 2018 Dec 14;20(12):e10165 [FREE Full text] [doi: 10.2196/10165] [Medline: 30552087]
51. Liu J, Zheng X, Zhang X, Feng Z, Song M, Lopez V. The evidence and future potential of WeChat in providing support for Chinese parents of pediatric patients undergoing herniorrhaphy. J Transcult Nurs 2020 Mar;31(2):114-120. [doi: 10.1177/1043659619841672] [Medline: 30983517]

52. Macken AP, Sasaki E, Quinn A, Cullen W, Leddin D, Dunne C, et al. Paediatric diabetes: information-seeking behaviours of families. Ir Med J 2014 Mar;107(3):87-88. [Medline: 24757896]
53. McCarty CA, Zatzick DF, Marcynyszyn LA, Wang J, Hilt R, Jinguji T, et al. Effect of collaborative care on persistent postconcussive symptoms in adolescents: a randomized clinical trial. JAMA Netw Open 2021 Feb 01;4(2):e210207 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0207] [Medline: 33635325]
54. McCarty C, Vander Stoep A, Violette H, Myers K. Interventions developed for psychiatric and behavioral treatment in the children's ADHD telemental health treatment study. J Child Fam Stud 2014 May 17;24(6):1735-1743. [doi: 10.1007/s10826-014-9977-5]
55. Meedya S, Win K, Yeatman H, Fahy K, Walton K, Burgess L, et al. Developing and testing a mobile application for breastfeeding support: the Milky Way application. Women Birth 2021 Mar;34(2):e196-e203. [doi: 10.1016/j.wombi.2020.02.006] [Medline: 32081557]
56. Price S, Ferisin S, Sharifi M, Steinberg D, Bennett G, Wolin KY, et al. Development and implementation of an interactive text messaging campaign to support behavior change in a childhood obesity randomized controlled trial. J Health Commun 2015;20(7):843-850. [doi: 10.1080/10810730.2015.1018582] [Medline: 25996181]
57. Ramelet AS, Fonjallaz B, Rio L, Zoni S, Ballabeni P, Rapin J, et al. Impact of a nurse led telephone intervention on satisfaction and health outcomes of children with inflammatory rheumatic diseases and their families: a crossover randomized clinical trial. BMC Pediatr 2017 Jul 17;17(1):168 [FREE Full text] [doi: 10.1186/s12887-017-0926-5] [Medline: 28716081]
58. Saidinejad M, Zorc J. Mobile and web-based education: delivering emergency department discharge and aftercare instructions. Pediatr Emerg Care 2014 Mar;30(3):211-216. [doi: 10.1097/PEC.0000000000000097] [Medline: 24589814]
59. Sharifi M, Dryden EM, Horan CM, Price S, Marshall R, Hacker K, et al. Leveraging text messaging and mobile technology to support pediatric obesity-related behavior change: a qualitative study using parent focus groups and interviews. J Med Internet Res 2013 Dec 06;15(12):e272 [FREE Full text] [doi: 10.2196/jmir.2780] [Medline: 24317406]
60. Spratling R, Faulkner MS, Chambers R, Lawrence P, Feinberg I, Hayat MJ. Establishing fidelity for the creating opportunities for personal empowerment: symptom and technology management resources (COPE-STAR) intervention. J Adv Nurs 2020 Dec;76(12):3440-3447 [FREE Full text] [doi: 10.1111/jan.14552] [Medline: 32989802]
61. Tutar Güven Ş, İşler Dalgiç A, Duman Ö. Evaluation of the efficiency of the web-based epilepsy education program (WEEP) for youth with epilepsy and parents: a randomized controlled trial. Epilepsy Behav 2020 Oct;111:107142. [doi: 10.1016/j.yebeh.2020.107142] [Medline: 32702651]
62. van der Gugten AC, Uiterwaal CS, Verheij TJ, van der Ent CK. E-health and consultation rates for respiratory illnesses in infants: a randomised clinical trial in primary care. Br J Gen Pract 2015 Feb;65(631):e61-e68 [FREE Full text] [doi: 10.3399/bjgp15X683485] [Medline: 25624309]
63. Blatz M, Dowling D, Underwood P, Bieda A, Graham G. A password-protected web site for mothers expressing milk for their preterm infants. Adv Neonatal Care 2017 Jun;17(3):222-229. [doi: 10.1097/ANC.0000000000000365] [Medline: 27902504]
64. Fiks AG, DuRivage N, Mayne SL, Finch S, Ross ME, Giacomini K, et al. Adoption of a portal for the primary care management of pediatric asthma: a mixed-methods implementation study. J Med Internet Res 2016 Jun 29;18(6):e172 [FREE Full text] [doi: 10.2196/jmir.5610] [Medline: 27357835]
65. Kobak KA, Stone WL, Wallace E, Warren Z, Swanson A, Robson K. A web-based tutorial for parents of young children with autism: results from a pilot study. Telemed J E Health 2011 Dec;17(10):804-808 [FREE Full text] [doi: 10.1089/tmj.2011.0060] [Medline: 22011005]
66. McGarry E, Vernon T, Baktha A. Brief report: a pilot online pivotal response treatment training program for parents of toddlers with autism spectrum disorder. J Autism Dev Disord 2020 Sep;50(9):3424-3431. [doi: 10.1007/s10803-019-04100-2] [Medline: 31190197]
67. Ruiz-Baqués A, Contreras-Porta J, Marques-Mejías M, Cárdenas Rebollo JM, Capel Torres F, Ariño Pla MN, et al. Evaluation of an online educational program for parents and caregivers of children with food allergies. J Investig Allergol Clin Immunol 2018;28(1):37-41 [FREE Full text] [doi: 10.18176/jiaci.0214] [Medline: 29461207]
68. Slater PJ, Fielden PE, Bradford NK. The oncology family app: providing information and support for families caring for their child with cancer. J Pediatr Oncol Nurs 2018;35(2):94-102. [doi: 10.1177/1043454217741874] [Medline: 29161974]
69. 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]
70. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008 May;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]
71. Schrauben SJ, Cavanaugh KL, Fagerlin A, Ikizler TA, Ricardo AC, Eneanya ND, et al. The relationship of disease-specific knowledge and health literacy with the uptake of self-care behaviors in CKD. Kidney Int Rep 2020 Jan;5(1):48-57 [FREE Full text] [doi: 10.1016/j.ekir.2019.10.004] [Medline: 31922060]

Abbreviations

JBI: Joanna Briggs Institute
MMAT: Mixed Methods Appraisal Tool
PROSPERO: International Prospective Register of Systematic Reviews
PRT: Pivotal Response Treatment

Edited by R Kukafka; submitted 30.06.21; peer-reviewed by M Fagnano, R Nightingale; comments to author 26.08.21; revised version received 06.10.21; accepted 29.10.21; published 22.12.21.

Please cite as:
Mörelius E, Robinson S, Arabiat D, Whitehead L
Digital Interventions to Improve Health Literacy Among Parents of Children Aged 0 to 12 Years With a Health Condition: Systematic Review
J Med Internet Res 2021;23(12):e31665
URL: https://www.jmir.org/2021/12/e31665
doi:10.2196/31665
PMID:34941559

Review

Value of the Electronic Medical Record for Hospital Care: Update From the Literature

Aykut Uslu[1], PhD; Jürgen Stausberg[2], MD, Prof Dr

[1]USLU Medizininformatik, Düsseldorf, Germany

[2]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Essen, Germany

Corresponding Author:
Jürgen Stausberg, MD, Prof Dr
Institute for Medical Informatics, Biometry and Epidemiology
University Hospital Essen
University Duisburg-Essen
Hufelandstrasse 55
Essen, 45122
Germany
Phone: 49 201 72377201
Fax: 49 201 72377333
Email: stausberg@ekmed.de

Related Articles:

Comment in: https://www.jmir.org/2022/3/e37419/

Comment in: https://www.jmir.org/2022/3/e37208/

Abstract

Background: Electronic records could improve quality and efficiency of health care. National and international bodies propagate this belief worldwide. However, the evidence base concerning the effects and advantages of electronic records is questionable. The outcome of health care systems is influenced by many components, making assertions about specific types of interventions difficult. Moreover, electronic records itself constitute a complex intervention offering several functions with possibly positive as well as negative effects on the outcome of health care systems.

Objective: The aim of this review is to summarize empirical studies about the value of electronic medical records (EMRs) for hospital care published between 2010 and spring 2019.

Methods: The authors adopted their method from a series of literature reviews. The literature search was performed on MEDLINE with "Medical Record System, Computerized" as the essential keyword. The selection process comprised 2 phases looking for a consent of both authors. Starting with 1345 references, 23 were finally included in the review. The evaluation combined a scoring of the studies' quality, a description of data sources in case of secondary data analyses, and a qualitative assessment of the publications' conclusions concerning the medical record's impact on quality and efficiency of health care.

Results: The majority of the studies stemmed from the United States (19/23, 83%). Mostly, the studies used publicly available data ("secondary data studies"; 17/23, 74%). A total of 18 studies analyzed the effect of an EMR on the quality of health care (78%), 16 the effect on the efficiency of health care (70%). The primary data studies achieved a mean score of 4.3 (SD 1.37; theoretical maximum 10); the secondary data studies a mean score of 7.1 (SD 1.26; theoretical maximum 9). From the primary data studies, 2 demonstrated a reduction of costs. There was not one study that failed to demonstrate a positive effect on the quality of health care. Overall, 9/16 respective studies showed a reduction of costs (56%); 14/18 studies showed an increase of health care quality (78%); the remaining 4 studies missed explicit information about the proposed positive effect.

Conclusions: This review revealed a clear evidence about the value of EMRs. In addition to an awesome majority of economic advantages, the review also showed improvements in quality of care by all respective studies. The use of secondary data studies has prevailed over primary data studies in the meantime. Future work could focus on specific aspects of electronic records to guide their implementation and operation.

(J Med Internet Res 2021;23(12):e26323) doi:10.2196/26323

KEYWORDS

cost analysis; costs and cost analyses; economic advantage; electronic medical records; electronic records; health care; hospitals; medical records systems computerized; quality of health care; secondary data

Introduction

This review is an update of 2 previous literature analyses on the benefits and costs of electronic medical records (EMRs), based on articles from 1966 to January 2004 [1] and from 2004 to 2010 [2]. Using the same method, this review explores the progress in evidence from empirical studies. The World Health Organization (WHO) has a clear position concerning the evidence for eHealth in general. Already in 2005, the WHO noted, "the potential impact that advances in information and communication technologies could have on health-care delivery..." [3]. Ten years later, the WHO put this straight by stating several advantages of electronic health records (EHRs) in the report of the third global survey on eHealth, which was produced by the Global Observatory for eHealth [4]:

- EHRs improve the quality, accuracy, and timeliness of patient information at the point of care.
- EHRs provide insights into health care costs, utilization, and outcomes.
- EHRs promote quality of care, reduce costs, support patient mobility, increase reliability of information, and provide access to patient information to multiple health care providers.
- Analyses from EHR data can highlight areas of concern and health services delivery.

The latter is emphasized in the current European digital strategy for data by creating a common European health data space that ensures interoperability of health data and in which every citizen has secure access to his or her EHR [5]. Consequently, many states adopted these visions and implemented national strategies for eHealth in general and for the EHR in particular (see [6] for an overview of Europe or [7] for country profiles from the Global Observatory for eHealth). In the United States, the meaningful use of health care information technology (IT) was fostered by the implementation of EHRs for all citizens until 2014 through the Health Information Technology for Economic and Clinical Health (HITECH) Act [8,9]. HITECH was successful, increasing the hospitals' adoption rate of a basic EHR from 9.4% in 2008 and 15.6% in 2010 to 97% in 2014 [10]. In Germany, the Patient Data Protection Act "obliges the public sickness funds to offer their clients an electronic patient record (EPR) not later than 1 January 2021" [11]. Furthermore, physician practices and hospitals are requested to support and to use the EPR based on the legal basis of an informed consent by the patients. In 2017, half of the German hospitals quoted the existence of an institutional electronic record similar to the situation in Austria [12]. Only the Swiss hospitals reported a higher proportion with 78%, a statistically significant difference to Germany.

EHRs will offer basic values by providing "the right information at the right time in the right place" [13]. This aim is achieved by improving the traditional function of patient records to store information relevant to the care. However, EHRs should additionally guide the process of clinical problem solving and should support clinical decision making [14]. In 1991, the Institute of Medicine (IOM) listed 4 ways to positively influence quality of care [14]: (1) improving quality of and access to clinical data, (2) integrating information over time and settings, (3) making knowledge available, and (4) providing decision support. Looking at costs, the IOM expected positive effects in 3 ways: (1) reducing unnecessary tests and services, (2) reducing administrative costs, and (3) increasing the productivity of health care professionals.

One might argue that a further discussion about the proposed value of an EMR is needless because of its nearly complete implementation. Nobody will vote for a fallback to paper. However, the implementation does not guarantee a positive perception by the users. In a recent survey including 208 physicians from 3 Norway hospitals [15], 72% of the physicians reported interrupted or delayed work at least once a week because the EHR hangs or crashes, and 53% of the physicians indicated that the EHR is cumbersome to use and adds to their workload. These results demonstrate a reasonable room for improvements, besides noncontroversial advantages that were reported in the study from Norway. Even if up-to-date health care cannot be imagined without an EMR, an ongoing evaluation of its advantages and disadvantages is a prerequisite for a well-considered further development and adjustment. In our sequence of literature reviews, we put the ultimate goals of health care in the middle, to provide a high level of care for reasonable costs in terms of effectiveness and efficiency [16]. Furthermore, the series of reviews allows a monitoring of the EMR's value over time by preserving the criteria for the selection and the appraisal of the included studies. The research questions were twofold. What is the effect of EMRs on the quality of inpatient care? What is the effect of EMRs on the costs for inpatient care?

Methods

Terminology of Electronic Records in Health Care

Concepts and terms denoting electronic records in health care are still not unambiguously defined [17]. Differences and similarities of "electronic medical records," "electronic patient records," and "electronic health records" are a matter of a long-lasting debate. In our reviews, we focused on electronic records used by health professionals and administrative staff for inpatient care, including, for example, physicians, nurses, radiologists, pharmacists, laboratory technicians, and radiographers [17]. Those records must not necessarily follow a patient lifelong. Therefore, we adopted the definition of an EMR by Waegemann [18]: an EMR is a "computer-stored collection of health information about a person, linked by a person identifier", with the application environment being a

hospital and including any care delivery being the full responsibility of the health care provider.

Search Strategy

The literature search was performed between March 10, 2019, and April 2, 2019, using MEDLINE. MEDLINE was accessed via PubMed [19]. The keyword "Medical Records Systems, Computerized" from the MeSH was separately combined with the MESH terms "technology assessment, biomedical", "costs and cost analysis", "health care costs", "cost savings", "cost effectiveness", "cost benefit", "cost analysis", "benefits and costs", "quality of health care", "outcome study", "outcome assessment, patient", and "critical care outcomes". Additionally, January 1, 2010, was defined as the earliest date of publication. After an exclusion of duplicates, interactive tutorials and reviews, and a restriction to the languages German and English, 1345 references remained.

Study Selection

Using titles and abstracts, both authors independently reviewed the 1345 literature references regarding the existence of an EMR, the application of an EMR in inpatient care, and an empirical analysis of benefits or costs. Explicitly excluded were studies in physician offices or about ambulatory care provided by hospitals, studies about picture archiving and communication systems, and studies about systems for computerized physician order entry (CPOE). The rating comprised the categories accept, refuse, and unclear. References rated as accept/accept and accept/unclear were qualified, references rated as refuse/refuse and refuse/unclear were rejected. References rated as accept/refuse or unclear/unclear were discussed and a final decision was reached based on a consensus. Herewith, 84 publications were qualified for the further evaluation (6.25%). From these, full texts of 79 papers could be obtained; for 5 papers, this was not possible. Textbox 1 shows the inclusion and exclusion criteria of both stages.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Acute care hospital
- Inpatient care
- Electronic medical record
- Empirical result
- Statement about costs
- Statement about benefits

Exclusion criteria

- Physician office
- Ambulatory care
- Picture archiving and communication system
- System for computerized physician order entry

Both authors again carried out the evaluation of the remaining 79 publications independently. This time, the evaluation was based on the full texts of the references. Both authors looked at concrete statements on benefits and costs, and gave a final recommendation about the inclusion into the review. References were finally included if they reached 2 or 3 positive votes from both authors (16/79 references, 20%). References were finally excluded if neither authors gave at least two positive votes (43/79 references, 54%). The remaining 20 references were discussed to reach a consensus about their inclusion for the review (25% from 79 references). Overall, the selection process produced 23 relevant studies that ultimately formed the subject of the detailed analysis, being 1.71% from the initially identified references (N=1345; Figure 1).

Interrater reliability during study selection was verified by calculating Cohen κ. In the first evaluation level based on titles and abstracts, the κ value was 0.185, indicating a slight agreement between the reviewers according to the interpretation of Landis and Koch [20] (Table 1). In the second evaluation level of full texts, the κ value was 0.428, indicating a moderate agreement. The interrater reliability was comparable to the previous reviews.

Figure 1. Selection and review process.

Table 1. Interpretation of κ values [20].

κ value	Level of agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Study Evaluation

For a semiquantitative evaluation of the studies, a catalog of criteria was drawn up focusing on the study design, the formal quality of the publication, the number of users included, the study duration, and the use of statistical tests. Each aspect of a study was rated 2, 1, or 0 points, with 2 being the best score for a study. Missing information was rated 0 points. The maximum number of points that could be achieved was therefore 10. In addition, the studies were described with regard to their origin, their application scenarios, and their target values. The approach proposed by Johnston et al [21] was adopted as basis for the evaluation method. The definition of the criteria was partly different between studies collecting primary data and studies analyzing existing, secondary data. The definition was carried out as described below.

Study Design

The assessment of the study design was based on the classification depicted in Table 2, according to Roine et al [22]. Different types of studies were ranked from 1 to 9 concerning the evidence hierarchy. The first stage, meta-analyses from randomized, controlled studies, was not a component of the inclusion criteria. The remaining study types were combined into the following 3 groups: randomized controlled studies (evidence stages 2 and 3); nonrandomized controlled studies (evidence stages 4, 5, 6, and 7); and uncontrolled clinical series, descriptive studies, consensus methods, application observations, and empirical reports (evidence stages 8 and 9). Studies in the first group received 2 points, studies in the second group 1 point, and the remaining studies 0 points. According to the proposal of Nathan and Gorman [23], all secondary data analyses were assigned to evidence stage 7 of Table 2 and were uniformly assigned 1 point. Therefore, the maximum number of points was reduced to 9 for those studies.

Table 2. Classification of study designs [22].

Evidence stage	Study design
1	Meta-analyses of randomized controlled trials
2	Large-sample randomized controlled trials
3	Small-sample randomized controlled trials
4	Nonrandomized controlled prospective studies
5	Nonrandomized controlled retrospective trials
6	Cohort studies
7	Case–control studies
8	Noncontrolled clinical series, descriptive studies, consensus methods
9	Anecdotes or case reports

Formal Quality of Publication

The publication should follow the international standard structure of scientific articles, that is, authors' names and affiliations on the title page, abstract, introduction, material and methods, results, discussion, conclusions, and references. For a publication in full compliance with this structure, 2 points were assigned; if the article provided a separate introduction and an explicit naming of authors and the medical environment, 1 point was assigned; otherwise, 0 points were given.

Number of Users

The number of EMR users can affect the reliability and the generalizability of the results. Therefore, 2 points were given for studies based on primary data with 20 or more users, 1 point for 6-19 users, and 0 points for less than 6 users or if no number of users was specified. For studies analyzing secondary data, the number of hospitals included was scored as follows: 2 points for a hospital number of 2000 and above; 1 point for a hospital number of 500 to 1999; and no points for a hospital number of 0-499 or missing information.

Implementation Duration

Primary data studies implemented for at least one year received 2 points, 1 point was given for a half to less than 1 year, and 0 points for less than a half year. For the secondary data studies, the evaluation periods were scored as follows: 2 points were awarded to a study for an evaluation period of 3 years or longer, 1 point for a period of 1 or 2 years, and no points for an implementation period of less than 1 year or in case of missing information.

Statistical Evaluation

Assessment and evaluation of scientific statements gain in evidential power with inferential statistical statements. Two points were given for studies reporting the result(s) of statistical analyses with full information concerning the level of significance, and 1 point for the description of a statistical test performed without indication of the level of significance. Otherwise, 0 point was given.

Results

Origins and Locations of the Studies

The 23 studies selected for the main evaluation [24-46] consisted of 6 primary and 17 secondary data studies (Table 3). Three (7, 18, and 23) of the 6 primary data studies were conducted in the United States (Wisconsin, North Carolina, and Massachusetts), 1 study each was conducted in China (21), Germany (3), and Japan (17). Sixteen secondary data studies originated from the United States, 1 from the Netherlands.

In the secondary data studies, 15 different data sources were used to analyze the issues of treatment quality, costs, and EMR equipment (Table 4). The most frequently used data sources stemmed from the American Hospital Association (AHA; 19 studies), the Healthcare Information and Management Systems Society (HIMSS; 11 studies), and the Centers for Medicare & Medicaid Services (CMS; 9 studies). They were followed by the Hospital Quality Alliance database (HQA), the National Database of Nursing Quality Indicators (NDNQI), and the Office of Statewide Health Planning and Development (OSHPD). The remaining sources were used only in 1 study.

Table 3. Characteristics of the included studies.

Study number	Reference	Country	Sample size	Period	Main outcomes
1	Adler-Milstein et al [24]	United States	191 hospitals	2 years	EHR[a] adoption is associated with better performance in terms of payment and length of stay in well-run institutions. EHR adoption may be associated with worse performance in poorly run institutions.
2	Adler-Milstein et al [25]	United States	2591 hospitals (2011)	4 years	Degree of EHR adoption is positively correlated with process adherence, patient satisfaction, and efficiency.
3	Castellanos et al [26]	Germany	Not indicated	6 years	Small increase in profit in the year after the introduction of the patient data management system.
4	DesRoches et al [27]	United States	3049 hospitals	6 months	Presence of clinical decision support is associated with small quality gains. No relationship between EHR level and overall risk-adjusted length of stay, risk-adjusted 30-day readmission rates, and risk-adjusted inpatient costs.
5	Elnahal et al [28]	United States	3101 hospitals	9 months	Higher rates of adoption of key EHR functions among high-quality hospitals.
6	Encinosa and Bae [29]	United States	2619 hospitals	1 year	EMRs[b] do not reduce the rate of patient safety events. In case of patient safety events, EMRs reduce deaths, readmissions, and spending.
7	Feblowitz et al [30]	United States	Not indicated	2 years	Length of stay increased after implementation of an electronic documentation. Mean time to disposition for admitted patients remained stable.
8	Furukawa et al [31]	United States	5066 hospitals	10 years	Advanced EMR applications may increase hospital costs and nurse staffing levels, as well as increase complications and decrease mortality for some conditions.
9	Furukawa et al [32]	United States	509 hospitals	5 years	Nurse-sensitive patient outcomes improved. EMR implementation may be associated with reduced demand for nurses.
10	Himmelstein et al [33]	United States	4000 hospitals	6 years	Hospital computerization has not achieved savings on clinical or administrative costs. More computerized hospitals might have a slight quality advantage for some conditions.
11	Jarvis et al [34]	United States	2988 hospitals	1 year	Most advanced EHRs have the greatest payoff in improving clinical process of care scores.
12	Jones et al [35]	United States	6057 hospitals	4 years	Availability of basic EHR is associated with a significant increase in health care quality for heart failure.
13	Joynt et al [36]	United States	1236 hospitals	4 years	Patients with stroke are more likely to receive guideline-driven components of care at hospitals with EHRs. Patients are slightly less likely to have a hospital stay longer than 4 days at hospitals with EHRs.
14	Kazley et al [37]	United States	1000 hospitals	1 year	In hospitals with advanced EHRs, patient costs are less compared with hospitals without advanced EHRs.
15	Lee et al [38]	United States	708 hospitals	8 years	Hospitals adopting EMRs experience shorter length of stay and lower 30-day mortality.
16	McCullough et al [39]	United States	3401 hospitals	4 years	Use of EHRs results in improvements in process-of-care measures for patients with heart failure or pneumonia.
17	Nakagawa et al [40]	Japan	Not indicated	7 years	EMR may decrease medical risks, but profitability does not rise more than the investments.

Study number	Reference	Country	Sample size	Period	Main outcomes
18	Schenarts et al [41]	United States	Not indicated	40 months	Implementation of the EMR is associated with an improvement in several complications and process measures.
19	Teufel et al [42]	United States	2307 hospitals	1 year	Advanced-stage EMR is associated with greater costs per case.
20	van Poelgeest et al [43]	Netherlands	67 hospitals	1 year	No statistically significant association between a hospital's EMR adoption and an overall quality or safety performance.
21	Xue et al [44]	China	251 physicians and 298,760 patient visits	5 years	Length of stay declines and mortality rate decreases with EMR. An EMR has no positive effect on patient costs.
22	Yanamadala et al [45]	United States	448,767 patients	1 year	Patients at hospitals with full EHR have the lowest rates of inpatient mortality, readmissions, and patient safety indicators.
23	Zlabek et al [46]	United States	Not indicated	Not indicated	Implementation of an inpatient EHR results in a rapid improvement in measures of cost of care.

[a]EHR: electronic health record.

[b]EMR: electronic medical record.

Table 4. Sources used by the secondary data studies.

Study number	Source (included years)			
	Quality	Costs	Electronic medical record	Other
1	AHA[a] (2009)	AHA (2009)	AHA IT Supplement (2009)	World Management Survey (2009)
2	AHA (2009-2012) CMS's[b] Hospital Compare (2009-2012)	CMS' EHR[c] Incentive Program reports (2009-2012)	AHA IT Supplement (2008-2011) CMS' EHR Incentive Program reports (2009-2012)	AHA annual survey (2008-2011)
4	AHA (2008) HQA[d] database (2009)	AHA (2008) Medicare Provider Analysis and Review (2006)	AHA IT Supplement (2008)	
5	HQA database (2006)		AHA IT Supplement (2009)	
6	MarketScan Commercial Claims and Encounter Database (2007) AHA (2007)	MarketScan Commercial Claims and Encounter Database (2007) AHA (2007)	AHA (2007)	
8	OSHPD[e] (1998-2007)	OSHPD (1998-2007)	HIMSS[f] (1998-2007)	OSHPD (1998-2007)
9	NDNQI[g] (2004-2008)	NDNQI (2004-2008)	HIMSS (2004-2008)	
10	Dartmouth Health Atlas (2008)	The Medicare Cost Reports	HIMSS (2003-2007)	
11	AHA (2008-2010)	CMS	HIMSS (2012?)	
12	AHA (2004-2007)		HIMSS (2003-2006)	
13	AHA (2007-2010)	AHA (2007-2010)	GWTG-Stroke[h] (2007-2010), linked with the AHA annual survey	
14		NIS[i] (2009)	HIMSS (2009)	
15	MEDPAR[j] (2000-2007)		HIMSS (2000-2007)	
16	AHA (2004-2007)	CMS (2004-2007)	HIMSS (2004-2007)	
19		HCUP KID[k] (2009)	HIMSS (2009)	
20	EMRAM[l] (2014)		EMRAM (2014)	
22	HCUP, SID[m] (2011)		AHA annual survey (2008, 2011)	

[a]AHA: American Hospital Association.

[b]CMS: Centers for Medicare & Medicaid Services.

[c]EHR: electronic health record.

[d]HQA: Hospital Quality Alliance database.

[e]OSHPD: Office of Statewide Health Planning and Development.

[f]HIMSS: Healthcare Information and Management Systems Society.

[g]NDNQI: National Database of Nursing Quality Indicators.

[h]GWTG-Stroke: Get With the Guidelines-Stroke.

[i]NIS: nursing information system.

[j]MEDPAR: Medicare Provider Analysis and Review

[k]HCUP KID: Healthcare Cost and Utilization Project Kids Inpatient Data.

[l]EMRAM: HIMSS Analytics EMR Adoption Model.

[m]SID: State Inpatient Databases.

Methodical Quality

The results of the semiqualitative assessment are presented in Table 5 and Multimedia Appendix 1. In the evaluation of the primary data studies, 2 (18 and 21) publications achieved a score of 6 points, 3 (3, 7, and 17) scored 4, and 1 (23) achieved only 2 points. No primary data study scored 0, 1, 3, 5, and 7-10 points. While in the secondary data studies 2 papers (2 and 10) achieved the maximal score of 9 points, another 4 (9, 13, 15, and 16) scored 8, 7 (4-6, 8, 11, 12, and 19) scored 7, 2 (1 and 14) scored 6, 1 (22) scored 5, and 1 (20) scored 4. No secondary data study scored 0-3 and 10 points. A total of 18 of the 23

studies scored 5 and more points (78%), while 5 remained below this score (22%). Only 2/6 (33%) primary data studies achieved 5 points or more. By contrast, 16/17 (94%) secondary data studies achieved a score of 5 points or more.

Two (1 and 19) of the primary data studies were randomized controlled trials; one (4) was a nonrandomized controlled trial; the remaining 3 belonged to a lower evidence stage. By definition, the 17 secondary studies were all assigned to evidence level 7. Fifteen (1-6, 9-11, 13-16, 19, and 20) studies followed the internationally accepted structure of scientific articles. The remaining 8 studies (7, 8, 12, 17, 18, and 21-23) lacked any formal structure.

Three (17, 18, and 21) of the 6 primary data studies had a user population of at least 20 or more. The remaining three (3, 7, and 23) did not provide any information. Five (3, 7, 17, 18, and 21) primary data studies had an implementation period of at least one year, 1 (23) less than 6 months. Eleven (2, 4-6, 8, 10-12, 16, 19, and 22) of the 17 secondary studies included at least 2000 hospitals, 4 (9, 13-15) 500 to less than 2000 hospitals, and 2 (1 and 20) less than 500 hospitals. Eight (2, 8-10, 12, 13, 15, and 16) of the secondary data studies analyzed data from at least three years, 1 (1) from 1 or 2 years, and 8 (4-6, 11, 14, 19, 20 and 22) from less than 1 year.

Nineteen (1, 2, 4-15, 18, 19, and 21-23) of the 23 studies supported the value of their results by statistical tests with full information on the level of significance. Two (16 and 20) studies stated that they had performed statistical tests but did not name them, and 2 (3 and 17) studies did not provide any information on them.

Table 5. Final score and conclusions of the included studies.

Study number	Reference	Data source	Final score	Cost reduction	Improvement in quality of care
1	Adler-Milstein et al [24]	S[a]	6	p[b]	p
2	Adler-Milstein et al [25]	S	9	p	p
3	Castellanos et al [26]	P[c]	4	p	n.a.[d]
4	DesRoches et al [27]	S	7	n	x[e]
5	Elnahal et al [28]	S	7	n.a.	p
6	Encinosa and Bae [29]	S	7	p	p
7	Feblowitz et al [30]	P	4	n[f]	x
8	Furukawa et al [31]	S	7	n	p
9	Furukawa et al [32]	S	8	p	p
10	Himmelstein et al [33]	S	9	n	x
11	Jarvis et al [34]	S	7	n.a.	p
12	Jones et al [35]	S	7	n.a.	p
13	Joynt et al [36]	S	8	p	p
14	Kazley et al [37]	S	6	p	n.a.
15	Lee et al [38]	S	8	p	p
16	McCullough et al [39]	S	8	n.a.	p
17	Nakagawa et al [40]	P	4	n	n.a.
18	Schenarts et al [41]	P	6	n.a.	p
19	Teufel et al [42]	S	7	n	n.a.
20	van Poelgeest et al [43]	S	4	n.a.	x
21	Xue et al [44]	P	6	n	p
22	Yanamadala et al [45]	S	5	n.a.	p
23	Zlabek et al [46]	P	2	p	n.a.

[a]S: secondary data studies.

[b]p: positive effect.

[c]P: primary data studies.

[d]n.a.: not assessed.

[e]x: positive effect without specific information.

[f]n: no positive effect.

Main Subjects

A total of 5 out of the 23 studies (3, 14, 17, 19, and 23) dealt solely with economic aspects of the use of an EMR, 7 (5, 11, 12, 16, 18, 20, and 22) dealt solely with the effects on the quality of care, and 11 studies (1, 2, 4, 6-10, 13, 15, and 21) dealt with both aspects (Tables 3 and 4). Primary data studies and secondary data studies were found in all groups. While 9 (39%) of the 23 studies (1-3, 6, 9, 13-15, and 23) showed an economically positive impact, 7 (30%) (4, 7, 8, 10, 17, 19, and 21) did not reveal monetary advantages due to the use of the EMR. Eighteen studies (1, 2, 4-13, 15, 16, 18, 20-22) looked at the impact of the use of an EMR on the quality of care. All of them (18/23 studies, 78%) found a positive effect. However, 4 (4, 7, 10, and 20) did not provide specific information about it. No study indicated evidence of disadvantages in the quality of treatment from the use of an EMR. Primary data studies and secondary data studies showed similar results.

One of the striking studies, Zlabek et al [46] looked at the effects of an EMR system on selected measures of cost of care and patient safety. They demonstrated the following outcomes (means and % change):

- Laboratory tests per week per hospitalization decreased from 13.9 to 11.4 (18).
- Radiology examinations per hospitalization decreased from 2.06 to 1.93 (6.3).
- Monthly transcription costs declined from US $74,596 to US $18,938 (74.6).
- Numbers of copy paper ordered per month decreased from 1668 to 1224 (26.6).
- Medication errors per 1000 hospital days decreased from 17.9 to 15.4 (14.0), while near misses per 1000 hospital days increased from 9.0 to 12.5 (38.9), and the percentage of medication events that were medication errors decreased from 66.5% to 55.2%.

In a national study about hospital computing and the costs and quality of care, Himmelstein et al [33] analyzed whether highly computerized hospitals had lower costs of care or administration, or better quality. They acquired the following outcomes in their work:

- Higher overall computerization scores correlated weakly with better quality scores for acute myocardial infarction, but not for heart failure, pneumonia, or the 3 conditions combined. In multivariate analyses, more computerized hospitals had a slightly better quality.
- Hospitals on the “Most Wired” list performed not better than others on quality, costs, or administrative costs.
- Hospitals’ administrative costs increased slightly but steadily, from 24.4% in 2003 to 24.9% in 2007. Higher administrative costs weakly predicted higher total Medicare spending, inpatient spending, and outpatient spending.

According to the study performed by Encinosa and Bae [29], many reforms in the Patient Protection and Affordable Care Act (ACA) underlie the use of EMRs to help contain costs. In this regard, the authors found that EMRs do not reduce the rate of patient safety events. However, once an event occurs, EMRs reduce death by 34%, readmissions by 39%, and spending by US $4850 (16%), a cost offset of US $1.75 per US $1 spent on IT capital. Thus, the authors concluded that EMRs contain costs by better coordinating care, a coordination that rescues patients from medical errors once they occur.

The study by Castellanos et al [26] analyzed cost and reimbursement data from a 25-bed intensive care unit at a German university hospital in a retrospective analysis, 3 years before and 3 years after the implementation of a patient data management system (PDMS). Costs and revenues increased continuously over the years. The profit of the investigated intensive care unit was fluctuating over the years and seemingly depending on other factors as well. They found a small increase in profit in the year after the introduction of the PDMS, but not in the following years. Therefore, a clear evidence for cost savings after the introduction of PDMS was not seen.

Discussion

Principal Findings

This review is an update of 2 previous analyses on the benefits and costs of EMRs, based on articles from 1966 to January 2004 [1] and 2004 to 2010 [2]. Using the same method, this review explored the progress in evidence from empirical studies. With a total of 19 of the 23 publications selected for evaluation (83%), studies from the United States dominated. Of the remaining 4 studies, 2 were conducted in Europe. Asia was represented by 1 Chinese and Japanese study each. South America, Africa, and Australia were not represented at all. Results of our reviews over the 3 periods showed a number of significant developments (Table 6). For example, the total number of initial hits had almost doubled. While the number of studies relevant to the evaluation remained more or less the same for the first and the current review, the second review produced almost one-third fewer studies. Remarkable in the current review was the predominant use of secondary data studies compared with primary data studies. In this context, highlighting the differences between primary and secondary studies should help to better assess the conclusions drawn from the results. While the primary data studies collected new and yet unexplored data, the secondary data studies used statistical processing of already existing data. In general, secondary data studies do not reach the evidence level of meta-analyses comprising also already existing but initially primary data. The most important advantage of primary data studies is that data can be collected and statistically evaluated in a targeted and problem-oriented manner. Their disadvantage is that specific surveys of patient data are often time-consuming and expensive compared with secondary data studies. Furthermore, in case of complex interventions, as it is the case for EMRs, primary data studies are often not feasible [47]. The advantage of secondary data studies is that comparatively few resources are required to prepare them. Their disadvantage is that the data were not collected specifically to answer the research questions as part of a specifically designed study design.

The annual number of studies on EMRs showed a continuous increase over our 3 review periods (Table 6). The same was true for the annual number of finally included studies. The methodological quality of the studies changed as well. While

only 35% of the studies scored more than 5 points in the first review (7/20), 74% of the studies scored more than 5 points in the third review (17/23). Among the finally included studies in the first review, costs were analyzed in 100% of the publications (20/20), with only 20% also focusing on quality of care (4/20). In the second review, both aspects were analyzed in 71% of the publications (5/7). In this review, costs were analyzed in 70% of the publications (16/23), quality of care in 78% (18/23).

The comparison of the 3 periods revealed a twofold shift. On the one hand, the studies' focus switched from an economical one to a clinical one. The percentage of studies concerned solely with costs decreased from 80% (16/20, 1966-2004) to 14% (1/7, 2002-2010) and 22% (5/23, 2010-2019). On the other hand, the positive effects of EMRs on quality of care became apparent over time. In the first review, none of the 4 studies concerned with quality of care presented well-defined advantages. In this review, this was the case in 14 of 18 studies analyzing the effects of EMRs on quality of care. The reasons for this shift remain speculative. The focus of EMRs might have changed from an administrative one to a patient-oriented one. Technological progress could have helped to achieve the clinical benefits that were an important motivator for the introduction of EMRs even in the early years [48]. In 1997, it was reported that costs remained a significant barrier for EHRs [49]. Now, experiences concerning the introduction, implementation, and an accompanying change management might have better prepared hospitals for the harvesting of clinical benefits and simultaneously for the limiting of additional costs.

Table 6. Number of studies considered for the reviews.

Review	Years, n	Hits without duplicates, n	Hits per year, mean	First selection, n	Finally included studies, n	Finally included studies per year, mean
First (1966-2004)	38	588	15.5	117	20	0.5
Second (2004-2010)	6	578	96.3	64	7	1.2
This (2010-2019)	9	1345	149.4	84	23	2.6

Limitations

The reliability between the 2 authors in selecting the papers was slight in the first phase (κ=0.185) and moderate in the second phase (κ=0.428). Both results were nearly equal compared with the 2 previous reviews, first phase 0.26 (review 1) and 0.192 (review 2), second phase 0.36 (review 1) and 0.399 (review 2). Unfortunately, measures of interrater reliability are usually not presented in systematic reviews. We assume that our results are not inferior in comparison to comparable reviews. The agreement was high in excluding references that do not fulfill the inclusion criteria. Differences occur in the detection of appropriate studies. To avoid the exclusion of false negatives, contrary votes and unclear votes were dissolved in a consensus. However, the extraction of the papers' main conclusions was a complex process. Misunderstandings and errors in this process cannot be completely ruled out. For example, authors' conclusions summarized in a paper's abstract could differ from individual results found in the paper's main text. The results of univariate and multivariate analyses may not agree and positive effects in one medical condition could be absent in another condition. Therefore, the review's rating is a pragmatical compromise to reach a meaningful conclusion.

The authors kept the EMR as type of intervention for all 3 reviews and attached great importance to an unaltered approach. This allowed the comparison of results over the whole series of reviews. The decision to maintain the focus on the EMR might be questioned because the literature addresses many different levels of IT used in hospitals. The results are therefore neither tailorable to more detailed types of IT providing only selective functionalities as CPOE nor generalizable to lifelong EPRs or to health information and communication technology overall. Nevertheless, through the clear and persistent focus, the authors gained reliable and valid conclusions beyond transitory trends and fashions.

Furthermore, the series maintained the same set of keywords. The authors could not rule out that newer functionalities of EMRs are not appropriately covered by this set. However, even then, the striking results supporting an indisputable positive effect of EMRs would be an underestimation of the actual situation. It is unlikely that newer functionalities decline the effects of EMRs on quality of care.

The detected studies represent primarily the perspectives of the United States and developed countries. Developed countries have the economic power to implement EMRs and to realize respective evaluation studies. This will not be the case for developing countries. However, the perspective for developing countries is similar. For example, Odekunle et al [50] reported for Sub-Saharan Africa the same vision as it was uncovered in our review. EHRs will improve quality of care in Sub-Saharan Africa, but high costs of procurement and maintenance of the EHR system hindered their widespread adoption until now.

Comparison With Prior Work

In 1963 the then American President, John F. Kennedy, was pointed to the potential of health record systems: "The application of computer technology to the recording, storage, and analysis of data collected in the course of observing and treating large numbers of ill people promises to advance our understanding of the cause, course, and control of disease" [51]. Forty-five years later, another American President (in 2009) proposed a fundamental change to the use of IT in the national health care system by passing the HITECH Act [52]. Besides other regulations, each person in the United States should have an EHR by 2014 [53]. With the idea of a meaningful use, health care providers and hospitals should be rewarded for using an EHR under the Medicare and Medicaid schedule. The time gap between expectations and routine application makes it clear that the proposed advantages were neither easy to demonstrate nor easy to achieve [54]. Even a proposal in 1991 for a nationwide

implementation of electronic records in the next decade failed [14]. Whether an evaluation of a technology in one country could be transferred to another one remains questionable, considering different health care systems and different strategies implemented with regard to the digitization of health care [55].

Our result of the positive impact of EMRs on the quality of care is supported by a systematic review by Campanella et al [56]. Their meta-analysis of 47 studies revealed a reduction of documentation time, a higher guideline adherence, and a lower number of medication errors and adverse drug events in the intervention group using an EHR. However, no association with mortality was found. Different to our review, the authors included studies on CPOE and did not focus on a specific area. The effect on mortality might be too small to be statistically significant even in a meta-analysis. Therefore, the inclusion of secondary data studies in our review series was reasonable. Thompson et al [57] also did not find a positive impact of EMRs on mortality. Besides, they did not find a positive impact on length of stay and costs. Their results were similar for record systems, CPOE, clinical decision support systems (CDSSs), and surveillance systems. In contrast to our results, Thompson et al [57] concluded that there "is not enough evidence to confidently state that electronic interventions have the ability to achieve the goal of improving quality and safety". Moja et al [58] also did not find effects of CDSSs on mortality in their meta-analysis based on 16 randomized controlled trials [58]. The authors stated, "most of the studies were underpowered and too short to prove or exclude an effect on mortality, and effects as large as a 25% increase or reduction could still be possible." In this day and age, where digitization is anywhere, it could become difficult to fill this gap with randomized controlled trials about EMRs using an appropriate control group. Besides secondary data analyses, ecological analysis might be worthwhile, even though the risk of an ecological fallacy exists [59]. With regard to CPOE as another subfunctionality of an EMR, Page et al [60] analyzed the evidence concerning a positive impact of quality of care. Defining a period overlapping with our study, 2000-2016, they included 23 studies with a control group. About half of the studies reported beneficial effects. However, the authors did not clearly distinguish between the effects of medication prescribing alerts as intervention and CPOE systems as infrastructure.

In summary, the impact of EMR subfunctionalities remains unclear in the literature. At a level beyond electronic records, the impact of health information exchange (HIE) as "the electronic transfer of patient data and health information between health care providers" is discussed [61]. Having EMRs as the condition, the exchange of data via HIE might bring the breakthrough in terms of quality of care and cost reduction. In their recent review, Sadoughi et al [62] considered 32 studies published between 2005 and 2016 that analyzed the financial or clinical impact of HIE. In that review, studies on EMRs were explicitly excluded. The majority of the studies were conducted in the United States (28/32), which is similar to our results. Furthermore, 19 studies were labeled as cohort studies, supporting our observation of a rather small number of controlled trials. Nearly all studies analyzing an improvement of quality showed a positive impact (16/17, 94%); 15/19 (79%) respective studies showed a positive effect on cost-effectiveness. With a similar span, these results from Sadoughi et al [62] match our review, with 78% of studies demonstrating an increase in quality of care and 56% demonstrating a reduction of costs. Contrary to a review including studies between 2003 and 2014 [63], Sadoughi et al [62] revealed a considerable progress in the use of HIE.

However, the advantages of EMRs have to be balanced with risks that are linked to IT not necessarily considered in evaluation studies. The relationship between the level of digitization and effects on quality and costs of care must not be linear. Higher levels of digitization might be correlated with higher risks that could lead to a reversion of the effect, as indicated by a study about the HITECH Act [64]. Therefore, it might be worthwhile to focus on the appropriate level of health IT instead of looking for global effects. Furthermore, the type of technology might not make the difference but rather the usability of the technology. For example, Roman et al [65] analyzed navigation-related issues in the field of EHRs. A lack in usability could induce risks for health care that lower the provided level of care. Finally, one should not forget that software, hardware, or electrical power supply can fail or can be a target for criminal attacks [66]. An overall perspective on the value of EMRs must therefore include a broader definition of assets and drawbacks.

Conclusions

Our literature review revealed a clear evidence about the value of EMRs. Only some primary data studies failed to demonstrate a reduction of costs after the implementation of an EMR. Quality of care improved in all respective studies. In comparison with our first review covering the period between 1996 and 2004, the picture changed completely. At that point, only 4 of 20 studies published benefits for the quality of care and 19 reported a reduction of costs. In parallel with the appearance of the first secondary data studies, the proportions turned around in the second review from 2004 to 2010. Interestingly, the positive effects on costs could not be completely confirmed by primary data studies now. To promote an extended use of EMRs, there must be a financial refund of additional costs, given the current scientific evidence. The switch from interventional studies to observational studies using publicly available data might have induced a bias in confirming everyday perceptions about electronic records in health care. Broader and better designed studies are needed to establish better scientific evidence regarding benefits of EMRs in hospital care. Nevertheless, further studies could focus on specific aspects of electronic records to guide their implementation and operation.

Acknowledgments

We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Included studies with scoring results.
[DOCX File , 19 KB - jmir_v23i12e26323_app1.docx]

References

1. Uslu AM, Stausberg J. Value of the electronic patient record: an analysis of the literature. J Biomed Inform 2008 Aug;41(4):675-682 [FREE Full text] [doi: 10.1016/j.jbi.2008.02.001] [Medline: 18359277]
2. Uslu A, Stausberg J. Value of the electronic medical record for hospital care: a review of the literature. Journal of Healthcare Engineering 2011 Sep;2(3):271-284. [doi: 10.1260/2040-2295.2.3.271]
3. World Health Organization. Resolutions and Decisions. WHA58.28 eHealth. URL: https://www.who.int/healthacademy/media/WHA58-28-en.pdf [accessed 2020-10-29]
4. World Health Organization. Global Observatory for eHealth. Global diffusion of eHealth: Making universal health coverage achievable. In: Report of the Third Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2016.
5. European C. A European strategy for data. Brussels, 19. URL: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf [accessed 2020-10-29]
6. Bonomi S. The electronic health record: a comparison of some European countries. In: Ricciardi F, Harfouche A, editors. Information and communication technologies in organizations and society. Past, present and future issues. Cham: Springer; 2016:33-50.
7. Global Observatory for eHealth. Atlas. URL: https://www.who.int/goe/publications/atlas/2015/ [accessed 2021-11-25]
8. Institute of Medicine (U.S.). Committee on Patient Safety and Health Information Technology. Health IT and patient safety: building safer systems for better care. Washington D.C: National Academy of Sciences; 2012.
9. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
10. Gold M, Mc L. Assessing HITECH implementation and lessons: 5 years later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
11. Federal Ministry of Health. Patientendaten-Schutz-Gesetz (PDSG) – Digitale Angebote schnell in die Versorgung bringen. Gesundheitsökonomie & Qualitätsmanagement 2020 Jul 13;25(03):149. [doi: 10.1055/a-1199-0942]
12. Naumann L, Esdar M, Ammenwerth E, Baumberger D, Hübner U. Same goals, yet different outcomes: analysing the current state of ehealth adoption and policies in Austria, Germany, and Switzerland using a mixed methods approach. Stud Health Technol Inform 2019 Aug 21;264:1012-1016. [doi: 10.3233/SHTI190377] [Medline: 31438077]
13. Tobias J, Stewart-Weeks M, editors. Could 9: connecting clinicians for better health care. Sydney: Cisco Systems/Intel Corporation; 2007.
14. Dick R, Steen E, editors. The computer-based patient record. An essential technology for health care. Washington, DC: National Academy Press; 1991.
15. Schopf TR, Nedrebø B, Hufthammer KO, Daphu IK, Lærum H. How well is the electronic health record supporting the clinical tasks of hospital physicians? A survey of physicians at three Norwegian hospitals. BMC Health Serv Res 2019 Dec 04;19(1):934 [FREE Full text] [doi: 10.1186/s12913-019-4763-0] [Medline: 31801518]
16. Das Deutsche Institut für Normung e.V. (DIN). Health care services - Quality management systems - Requirements based on EN ISO 9001:2008; German version EN 15224:2012. Berlin, Germany: Beuth; 2012.
17. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
18. Waegemann CP. The five levels of electronic health records. MD Comput 1996;13(3):199-203. [Medline: 8935995]
19. PubMed. URL: http://www.pubmed.gov/ [accessed 2021-11-25]
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
21. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. Ann Intern Med 1994 Jan 15;120(2):135-142. [doi: 10.7326/0003-4819-120-2-199401150-00007] [Medline: 8256973]
22. Roine R, Ohinmaa A, Hailey D. Assessing telemedicine: a systematic review of the literature. CMAJ 2001 Sep 18;165(6):765-771 [FREE Full text] [Medline: 11584564]
23. Nathan P, Gormann J, editors. A Guide to Treatments That Work (Fourth Edition). Oxford, UK: Oxford University Press; 2015.
24. Adler-Milstein J, Woody Scott K, Jha A. Leveraging EHRs to improve hospital performance: the role of management. Am J Manag Care 2014 Nov;20(11 Spec No. 17):SP511-SP519 [FREE Full text] [Medline: 25811825]

25. Adler-Milstein J, Everson J, Lee SD. EHR adoption and hospital performance: time-related effects. Health Serv Res 2015 Dec 16;50(6):1751-1771 [FREE Full text] [doi: 10.1111/1475-6773.12406] [Medline: 26473506]
26. Castellanos I, Schüttler J, Prokosch H, Bürkle T. Does introduction of a Patient Data Management System (PDMS) improve the financial situation of an intensive care unit? BMC Med Inform Decis Mak 2013 Sep 16;13(1):107 [FREE Full text] [doi: 10.1186/1472-6947-13-107] [Medline: 24041117]
27. DesRoches CM, Campbell EG, Vogeli C, Zheng J, Rao SR, Shields AE, et al. Electronic health records' limited successes suggest more targeted uses. Health Affairs 2010 Apr;29(4):639-646. [doi: 10.1377/hlthaff.2009.1086] [Medline: 20368593]
28. Elnahal SM, Joynt KE, Bristol SJ, Jha AK. Electronic health record functions differ between best and worst hospitals. Am J Manag Care 2011 Apr;17(4):e121-e147 [FREE Full text] [Medline: 21774097]
29. Encinosa WE, Bae J. Health information technology and its effects on hospital costs, outcomes, and patient safety. Inquiry 2011 Dec 01;48(4):288-303 [FREE Full text] [doi: 10.5034/inquiryjrnl_48.04.02] [Medline: 22397059]
30. Feblowitz J, Takhar SS, Ward MJ, Ribeira R, Landman AB. A custom-developed emergency department provider electronic documentation system reduces operational efficiency. Ann Emerg Med 2017 Nov;70(5):674-682.e1 [FREE Full text] [doi: 10.1016/j.annemergmed.2017.05.032] [Medline: 28712608]
31. Furukawa M, Raghu T, Shao BBM. Electronic medical records, nurse staffing, and nurse-sensitive patient outcomes: evidence from California hospitals, 1998-2007. Health Serv Res 2010 Aug;45(4):941-962 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01110.x] [Medline: 20403065]
32. Furukawa MF, Raghu T, Shao BBM. Electronic medical records, nurse staffing, and nurse-sensitive patient outcomes: evidence from the national database of nursing quality indicators. Med Care Res Rev 2011 Jun 11;68(3):311-331. [doi: 10.1177/1077558710384877] [Medline: 21075750]
33. Himmelstein DU, Wright A, Woolhandler S. Hospital computing and the costs and quality of care: a national study. Am J Med 2010 Jan;123(1):40-46. [doi: 10.1016/j.amjmed.2009.09.004] [Medline: 19939343]
34. Jarvis B, Johnson T, Butler P, O'Shaughnessy K, Fullam F, Tran L, et al. Assessing the impact of electronic health records as an enabler of hospital quality and patient satisfaction. Academic Medicine 2013;88(10):1471-1477. [doi: 10.1097/acm.0b013e3182a36cab]
35. Jones S, Adams J, Schneider E, Ringel J, McGlynn E. Electronic health record adoption and quality improvement in US hospitals. Am J Manag Care 2010 Dec;16(12 Suppl HIT):SP64-SP71 [FREE Full text] [Medline: 21314225]
36. Joynt KE, Bhatt DL, Schwamm LH, Xian Y, Heidenreich PA, Fonarow GC, et al. Lack of impact of electronic health records on quality of care and outcomes for ischemic stroke. J Am Coll Cardiol 2015 May 12;65(18):1964-1972 [FREE Full text] [doi: 10.1016/j.jacc.2015.02.059] [Medline: 25953748]
37. Kazley AS, Simpson AN, Simpson KN, Teufel R. Association of electronic health records with cost savings in a national sample. Am J Manag Care 2014 Jun 01;20(6):e183-e190 [FREE Full text] [Medline: 25180501]
38. Lee J, Kuo Y, Goodwin JS. The effect of electronic medical record adoption on outcomes in US hospitals. BMC Health Serv Res 2013 Mar 01;13:39 [FREE Full text] [doi: 10.1186/1472-6963-13-39] [Medline: 23375071]
39. McCullough JS, Casey M, Moscovice I, Prasad S. The effect of health information technology on quality in U.S. hospitals. Health Aff (Millwood) 2010 Apr;29(4):647-654. [doi: 10.1377/hlthaff.2010.0155] [Medline: 20368594]
40. Nakagawa Y, Tomita N, Irisa K, Yoshihara H, Nakagawa Y. Management evaluation about introduction of electric medical record in the national hospital organization. Stud Health Technol Inform 2013;192:1000. [Medline: 23920774]
41. Schenarts PJ, Goettler CE, White MA, Waibel BH. An objective study of the impact of the electronic medical record on outcomes in trauma patients. The American Surgeon 2012 Nov 01;78(11):1249-1254. [doi: 10.1177/000313481207801134]
42. Teufel RJ, Kazley AS, Ebeling MD, Basco WT. Hospital electronic medical record use and cost of inpatient pediatric care. Acad Pediatr 2012;12(5):429-435. [doi: 10.1016/j.acap.2012.06.004] [Medline: 22819201]
43. van Poelgeest R, Heida J, Pettit L, de Leeuw RJ, Schrijvers G. The association between ehealth capabilities and the quality and safety of health care in the Netherlands: comparison of HIMSS Analytics EMRAM data with Elsevier's 'The Best Hospitals' data. J Med Syst 2015 Sep 5;39(9):90 [FREE Full text] [doi: 10.1007/s10916-015-0274-7] [Medline: 26242750]
44. Xue Y, Liang H, Wu X, Gong H, Li B, Zhang Y. Effects of electronic medical record in a Chinese hospital: a time series study. Int J Med Inform 2012 Oct;81(10):683-689. [doi: 10.1016/j.ijmedinf.2012.05.017] [Medline: 22727614]
45. Yanamadala S, Morrison D, Curtin C, McDonald K, Hernandez-Boussard T. Electronic health records and quality of care: an observational study modeling impact on mortality, readmissions, and complications. Medicine (Baltimore) 2016 May;95(19):e3332 [FREE Full text] [doi: 10.1097/MD.0000000000003332] [Medline: 27175631]
46. Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. J Am Med Inform Assoc 2011 Mar 01;18(2):169-172 [FREE Full text] [doi: 10.1136/jamia.2010.007229] [Medline: 21292703]
47. Campbell M, Fitzpatrick R, Haines A, Kinmonth A, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
48. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. New York, NY: Springer; 1991.
49. Dick RS, Steen EB, Detmer DE. The Computer-Based Patient Record. An Essential Technology for Health Care (Revised Edition). Washington, DC: National Academy Press; 1997.

50. Odekunle FF, Srinivasan S, Odekunle RO. Why Sub-Saharan Africa lags in electronic health record (EHR) adoption and possible strategies to increase EHR adoption in this region. Journal of Health Informatics in Africa 2018;5:8-15. [doi: 10.12856/JHIA-2018-v5-i1-147]
51. Presidents Science Advisory Comittee Life Science Panel. Some New Technologies and Their Promise for the Life Sciences. Washington, DC: US Government Printing Office; 1963. URL: https://www.cia.gov/readingroom/docs/CIA-RDP66R00546R000200100002-1.pdf [accessed 2021-11-29]
52. Ohno-Machado L. Electronic health records: monitoring the return on large investments. J Am Med Inform Assoc 2013 Jun 01;20(e1):e1-e1 [FREE Full text] [doi: 10.1136/amiajnl-2013-001966] [Medline: 23687172]
53. One Hundred Eleventh Congress of the United States of America. American Recovery and Reinvestment Act of 2009. Washington, DC: Congress of the United States of America; 2009. URL: https://www.fcc.gov/general/american-recovery-and-reinvestment-act-2009 [accessed 2021-11-29]
54. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2018 Mar 06;25(S 01):S48-S61. [doi: 10.15265/iys-2016-s006]
55. Wilson K, Khansa L. Migrating to electronic health record systems: A comparative study between the United States and the United Kingdom. Health Policy 2018 Nov;122(11):1232-1239. [doi: 10.1016/j.healthpol.2018.08.013] [Medline: 30193980]
56. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Mar 30;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
57. Thompson G, O'Horo JC, Pickering BW, Herasevich V. Impact of the electronic medical record on mortality, length of stay, and cost in the hospital and ICU: a systematic review and metaanalysis. Crit Care Med 2015 Jun;43(6):1276-1282. [doi: 10.1097/CCM.0000000000000948] [Medline: 25756413]
58. Moja L, Kwag KH, Lytras T, Bertizzolo L, Brandt L, Pecoraro V, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. Am J Public Health 2014 Dec;104(12):e12-e22. [doi: 10.2105/ajph.2014.302164]
59. Morgenstern H. Uses of ecologic analysis in epidemiologic research. Am J Public Health 1982 Dec;72(12):1336-1344. [doi: 10.2105/ajph.72.12.1336] [Medline: 7137430]
60. Page N, Baysari MT, Westbrook JI. A systematic review of the effectiveness of interruptive medication prescribing alerts in hospital CPOE systems to change prescriber behavior and improve patient safety. Int J Med Inform 2017 Sep;105:22-30. [doi: 10.1016/j.ijmedinf.2017.05.011] [Medline: 28750908]
61. Esmaeilzadeh P, Sambasivan M. Health Information Exchange (HIE): A literature review, assimilation pattern and a proposed classification for a new policy approach. J Biomed Inform 2016 Dec;64:74-86 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.011] [Medline: 27645322]
62. Sadoughi F, Nasiri S, Ahmadi H. The impact of health information exchange on healthcare quality and cost-effectiveness: A systematic literature review. Comput Methods Programs Biomed 2018 Jul;161:209-232. [doi: 10.1016/j.cmpb.2018.04.023] [Medline: 29852963]
63. Rudin RS, Motala A, Goldzweig CL, Shekelle PG. Usage and effect of health information exchange. Ann Intern Med 2014 Dec 02;161(11):803. [doi: 10.7326/m14-0877]
64. Appari A, Eric Johnson M, Anthony DL. Meaningful use of electronic health record systems and process quality of care: evidence from a panel data analysis of U.S. acute-care hospitals. Health Serv Res 2013 Apr;48(2 Pt 1):354-375 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01448.x] [Medline: 22816527]
65. Roman LC, Ancker JS, Johnson SB, Senathirajah Y. Navigation in the electronic health record: A review of the safety and usability literature. J Biomed Inform 2017 Mar;67:69-79 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.005] [Medline: 28088527]
66. Sax U, Lipprandt M, Röhrig R. The rising frequency of IT blackouts indicates the increasing relevance of IT emergency concepts to ensure patient safety. Yearb Med Inform 2018 Mar 06;25(01):130-137. [doi: 10.15265/iy-2016-038]

Abbreviations

ACA: Affordable Care Act
AHA: American Hospital Association
CDSS: computer decision support systems
CMS: Centers for Medicare & Medicaid Services
CPOE: computerized physician order entry
EHR: electronic health record
EMR: electronic medical record
EPR: electronic patient record
GWTG-Stroke: Get With the Guidelines-Stroke
HCUP KID: Healthcare Cost and Utilization Project Kids Inpatient Data

HIE: health information exchange
HIMSS: Healthcare Information and Management Systems Society
HITECH: Health Information Technology for Economic and Clinical Health
HQA: Hospital Quality Alliance database
IOM: Institute of Medicine
IT: information technology
NDNQI: National Database of Nursing Quality Indicators
NIS: nursing information system
OSHPD: Office of Statewide Health Planning and Development
PDMS: patient data management system
RCT: randomized controlled trial
SID: State Inpatient Databases
WHO: World Health Organization

Edited by R Kukafka; submitted 07.12.20; peer-reviewed by C Gibson, W Sermeus; comments to author 26.03.21; revised version received 27.04.21; accepted 08.10.21; published 23.12.21.

Please cite as:
Uslu A, Stausberg J
Value of the Electronic Medical Record for Hospital Care: Update From the Literature
J Med Internet Res 2021;23(12):e26323
URL: https://www.jmir.org/2021/12/e26323
doi:10.2196/26323
PMID:34941544

Review

Methods to Generate Innovative Research Ideas and Improve Patient and Public Involvement in Modern Epidemiological Research: Review, Patient Viewpoint, and Guidelines for Implementation of a Digital Cohort Study

Gloria A Aguayo[1], MD, PhD; Catherine Goetzinger[1], MSc; Renza Scibilia[2,3]; Aurélie Fischer[1], MSc; Till Seuring[4], PhD; Viet-Thi Tran[5,6], MD, PhD; Philippe Ravaud[5,6], MD, PhD; Tamás Bereczky[7], PhD; Laetitia Huiart[1], MD, PhD; Guy Fagherazzi[1], PhD

[1]Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

[2]Diabetes Australia, Melbourne, Australia

[3]Diabetogenic, Melbourne, Australia

[4]Luxembourg Institute of Socio-Economic Research, Esch/Alzette, Luxembourg

[5]Centre of Research in Epidemiology and Statistic Sorbonne Paris Cité, National Institute of Health and Medical Research (INSERM), French National Institute for Agricultural Research (INRA), Université de Paris, Paris, France

[6]Centre d'Epidémiologie Clinique, Hôpital Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, Paris, France

[7]European Patients' Academy on Therapeutic Innovation, Brussels, Belgium

Corresponding Author:
Gloria A Aguayo, MD, PhD
Deep Digital Phenotyping Research Unit, Department of Population Health
Luxembourg Institute of Health
1 AB rue Thomas Edison
Strassen, 1445
Luxembourg
Phone: 352 26970770
Fax: 352 269770719
Email: gloria.aguayo@lih.lu

Abstract

Background: Patient and public involvement (PPI) in research aims to increase the quality and relevance of research by incorporating the perspective of those ultimately affected by the research. Despite these potential benefits, PPI is rarely included in epidemiology protocols.

Objective: The aim of this study is to provide an overview of methods used for PPI and offer practical recommendations for its efficient implementation in epidemiological research.

Methods: We conducted a review on PPI methods. We mirrored it with a patient advocate's viewpoint about PPI. We then identified key steps to optimize PPI in epidemiological research based on our review and the viewpoint of the patient advocate, taking into account the identification of barriers to, and facilitators of, PPI. From these, we provided practical recommendations to launch a patient-centered cohort study. We used the implementation of a new digital cohort study as an exemplary use case.

Results: We analyzed data from 97 studies, of which 58 (60%) were performed in the United Kingdom. The most common methods were workshops (47/97, 48%); surveys (33/97, 34%); meetings, events, or conferences (28/97, 29%); focus groups (25/97, 26%); interviews (23/97, 24%); consensus techniques (8/97, 8%); James Lind Alliance consensus technique (7/97, 7%); social media analysis (6/97, 6%); and experience-based co-design (3/97, 3%). The viewpoint of a patient advocate showed a strong interest in participating in research. The most usual PPI modalities were research ideas (60/97, 62%), co-design (42/97, 43%), defining priorities (31/97, 32%), and participation in data analysis (25/97, 26%). We identified 9 general recommendations and 32 key PPI-related steps that can serve as guidelines to increase the relevance of epidemiological studies.

Conclusions: PPI is a project within a project that contributes to improving knowledge and increasing the relevance of research. PPI methods are mainly used for idea generation. On the basis of our review and case study, we recommend that PPI be included

at an early stage and throughout the research cycle and that methods be combined for generation of new ideas. For e-cohorts, the use of digital tools is essential to scale up PPI. We encourage investigators to rely on our practical recommendations to extend PPI in future epidemiological studies.

(J Med Internet Res 2021;23(12):e25743) doi:10.2196/25743

KEYWORDS

patient and public involvement; workshops; surveys; focus groups; co-design; digital cohort study; digital epidemiology; social media; mobile phone

Introduction

Background

Patient and public involvement (PPI) in research is defined as "research being carried out 'with' or 'by' members of the public rather than 'to,' 'about,' or 'for' them" [1]. PPI means that patients or the public are actively involved in the research process rather than being included only as participants. Public engagement is the 2-way process of engagement activities that benefits both the researcher and the public [2].

Research that involves patients and the public can reduce the mismatch between what matters to patients and what is actually being done in the research [3]. A waste of research resources can be generated when the needs of those likely to use the research results, such as patients and caregivers, are not taken into account [4]. PPI can contribute to the identification and selection of high-priority research questions, planning and performing of more focused research, and improvement of participants' enrollment in clinical trials [5]. Ultimately, this can result in a higher societal benefit through better use of resources for research. PPI improves the quality of the study and makes research more relevant [6].

Involve, a UK-funded program, aims to improve the quality of research through the integration of PPI throughout the research cycle (identification and prioritization, commissioning, design and management, implementation, dissemination, implementation, and evaluation). Involve has published a report with guidelines to help researchers start new projects when they intend to include PPI in their projects [7].

Digital epidemiology has the same objectives as epidemiology, which are the observation of disease patterns, their evolution, and the causes of these patterns to improve population health and prevent diseases, but digital epidemiology uses digital data [8]. A digital or e-cohort study can integrate data that were not generated for the research (social media and registries) or were generated with digital tools (wearables, sensors, smartphone technologies, and e-questionnaires through web platforms) [9]. Before starting any epidemiological, clinical, or population-based study, researchers need to choose the best methodology to incorporate PPI throughout the project [10]. However, although there are approaches to integrate PPI in a research project, we think that there are no clear recommendations of which methods are the most appropriate, in particular with respect to the launch of cohort studies with digital sources of data.

Although PPI is recommended in research projects, this involvement is often not described or is incompletely reported [11]. A reason for underreporting may be to avoid describing an unsuccessful PPI attempt or that there was no involvement [12]. In addition, there is some evidence that PPI is seldom used in many countries [13]. In the case of cohort studies and, in particular, e-cohorts, we believe that this insufficient involvement may be due to a lack of knowledge of the methods, barriers, or facilitators to apply PPI. In addition, we think that there is a need to have concrete and clear examples for applying PPI in this type of study.

Objectives

The aims of this study are to (1) review methodologies used to include PPI in research, (2) provide the viewpoint of a final user of research results, and (3) provide practical guidelines and recommendations about how to initiate and run an e-cohort study with PPI based on the review and the point of view of a patient advocate.

Methods

This work entails 3 parts: a narrative review about methodology and description of PPI, a viewpoint of a patient advocate, and a case study with practical guidelines and recommendations, illustrated by the implementation of a digital cohort study.

Review

Data for this review were identified by searches of PubMed, Google Scholar, targeted websites about PPI, reports, and existing PPI guidelines, as well as Google Search and references from relevant studies. We used the following search terms: "patient and public participation," "patient engagement," "patient involvement," "consumer involvement," "community involvement," "participatory health research," "community based research," "research ideas," "co-writing," "coproduction," "co-design," "cohort study," "e-cohort," and "longitudinal study." We included original studies describing PPI methods using the Involve definition with at least one type of PPI in the study. We included information on studies published in English from 2000 to the present.

Viewpoint of a Patient Advocate

We used relevant definitions from the European Patients' Academy on Therapeutic Innovation for patients. We defined patients as individuals with personal experience of the disease, caregivers as individuals supporting a patient, and patient advocates as individuals representing large numbers of patients with a specific disease [14].

We invited a patient advocate to present her perspective and expectations regarding PPI in the context of diabetes research

and the use of digital tools and, in particular, the use of social media.

Recommendations

We integrated the results of the review and the patient advocate's viewpoint to identify practical guidelines on how to increase PPI in future epidemiological studies. We used the implementation of a digital cohort study as an exemplary case for testing and illustrating established guidelines for PPI [15]. In addition, we integrated in our recommendations the revised version of Guidance for Reporting Involvement of Patients and the Public (GRIPP2) checklist as an instrument to improve the quality of PPI reporting [16].

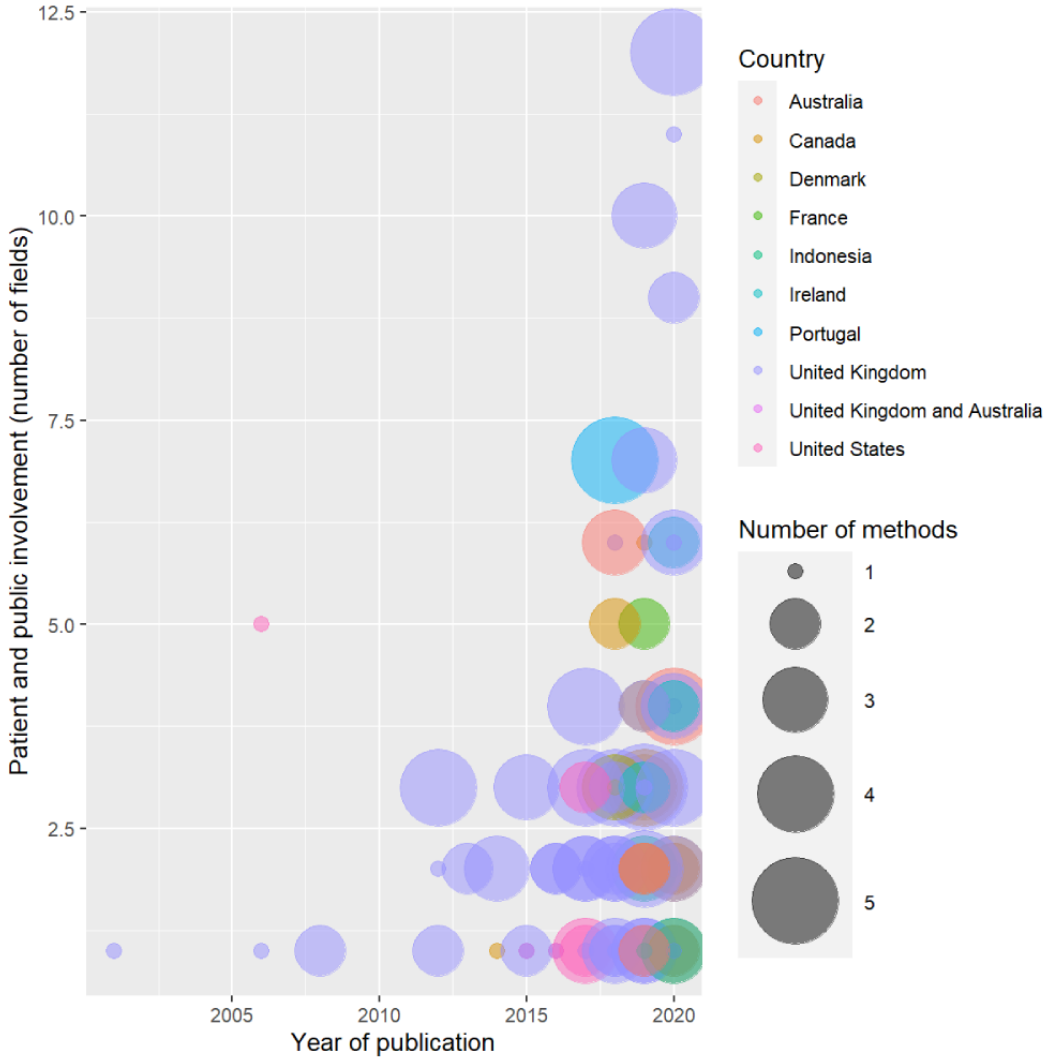
This study was based on a collaboration of patients and researchers. A patient advocate (RS), patient researchers (TS and TB), and researchers (GAA, CG, AF, VTT, PR, LH, and GF) were involved in the preparation of the manuscript (cowriting, editing, and critical review).

Results

Review of PPI Methods

We analyzed data from 97 studies published from 2000 to 2020. The studies were performed in 9 countries: the United Kingdom (58/97, 60%), Canada (13/97, 13%), the United States (8/97, 8%), Australia (6/97, 6%), Ireland (6/97, 6%), Denmark (2/97, 2%), France (1/97, 1%), Portugal (1/97, 1%), Indonesia (1/97, 1%), and the United Kingdom and Australia (1/97, 1%). The most frequent methods of PPI were workshops (47/97, 48%); surveys (33/97, 34%); meetings, events, or conferences (28/97, 29%); focus groups (25/97, 26%); interviews (23/97, 24%); consensus techniques (8/97, 8%); James Lind Alliance consensus technique (7/97, 7%); social media analysis (6/97, 6%); and experience-based co-design (3/97, 3%). Of the 97 studies, 34 (35%) used only 1 method, whereas 30 (31%), 22 (23%), 8 (8%), and 3 (3%) used 2, 3, 4, and 5 methods, respectively. The use of ≥3 methods together was observed from 2017 onwards (Figure 1 and Multimedia Appendix 1 [17-114]).

Figure 1. Number of fields or areas (ie, 1 field=involved in research ideas and 2 fields=research ideas and co-design) in which patients, carers, or the public were involved (y-axis); number of methods (circles); and countries (colors) where the studies were performed from 2000 to 2020 (x-axis). Patient and public involvement increases over time and at different stages of involvement. The size of each circle represents the number of methods used for patient and public involvement. Circles representing a combination of methods are very common in recently published studies. The most represented country is the United Kingdom.



Format of Involvement

Broad Definition of PPI Used

When considering methods of approaching PPI, 2 situations should be distinguished. The first situation is where patients or the public provide data on ideas or research priorities. The second situation is where patients or the public are actively involved in decisions about the research and perform or collaborate in the design, data analysis, data interpretation, writing, or diffusion of results. For this review, we used a broad definition of PPI that includes both situations: patients or the public involved in generating research ideas and patients or the public actively involved in the research cycle [115].

Workshops

We found that the most frequent method for approaching PPI was a workshop. Workshops are group activities where participants discuss a defined topic and decide actions. There are different ways of organizing a workshop. The first way is to organize it always in the same place and at the same hour. Morris et al [116] emphasized the importance of having enough space and time for the discussion. In addition, the number of participants by workshop or by discussion group may be limited. Mackintosh et al [17] proposed separate groups of users and service providers of 3-4 people. In contrast, Kelemen et al [18] organized full-day workshops with 20 to 25 participants of different origins and ages and divided into 2 groups randomly assigned. Patients prefer places with quiet rooms and parking lots, with meeting and meal times that allow them to participate and manage their medications [19]. An alternative to a workshop held at a specific location is to organize mobile workshops to facilitate the participation of people with busy schedules or mobility issues. Eccles et al [20] organized this type of workshop with a maximum of 10 participants and a short duration (30 minutes). These workshops were usually offered on the sites of existing organizations, which improved the success rate [20]. When face-to-face workshops are not possible, a workshop by videoconference can be organized [21]. Workshops can also be organized as consensus meetings [15,22], with a second-round workshop [117] or by applying other techniques to improve explicit involvement, such as experience-based co-design [23]. Workshops are usually analyzed with qualitative research methods such as thematic analysis [118] (Multimedia Appendix 1).

Surveys

Surveys aim to obtain information in an easy, individual, and feasible way. They can include open-ended questions. For example, a citizen science study asked participants the *magic wand* question ("If you had a magic wand, what would you change in your healthcare?") [24]. Surveys can also include closed-ended questions. Little et al [25] performed a study to identify patients' preferences for a patient-centered consultation using Likert scale questions that were previously tested and validated in a pilot study. To respond to a survey, patients and the public could be contacted in person [25], through a web-based survey [119], or by telephone [120]. Several studies included a second round of questions for improving ideas or establishing priorities [26,121].

Focus Groups

Focus groups are group activities, a form of qualitative research that includes people with certain similarities such as a therapeutic area or illness or demographic or socioeconomic group to discuss their opinions and beliefs about a topic familiar to the participants [27]. An advantage of focus groups is the potential to obtain multiple opinions and different understandings of the same question, which can enrich explanations for a certain issue. A disadvantage is that focus groups require a trained interviewer to conduct them, and interaction among participants can influence the outcome [122].

Meetings

Meetings are group activities where patients, carers, caregivers, and researchers participate in the project discussing and co-organizing activities for PPI [15]. They are complementary to workshops or focus groups and are also used to prepare these activities and to debrief methods or actions for PPI [28]. Meetings can be organized as festivals and public events, conferences [29], or web-based activities [30].

Semistructured Interviews

Semistructured interviews are a qualitative research method where the researcher discusses a defined topic individually and informally with the participant [123]. Using this method, nonresponse or direct behavioral observations, individual reflections on a specific research question, can be incorporated as outcomes; the advantage is that the participant is not influenced by other persons [124,125].

Consensus Techniques

Consensus techniques analyze a project's chances of success by bringing together a group of experts at a workshop, meeting, or conference to discuss solutions and reach agreements [31]. James Lind Alliance is a consensus technique bringing clinicians, patients, carers, and the public together to achieve the convergence of opinions for establishing research priorities. The process starts with a web-based survey to collect research ideas from the public. Next, only unanswered research questions are selected. With a second web-based survey, the public prioritizes the selected questions. At the end of the process, the steering committee chooses the top 10 research priorities [126].

Persona-Scenario

Persona-scenario is a method for co-design in PPI where a fictitious user (persona) is created to communicate in a committed manner. Next, a scenario is proposed based on a story with an actor (the persona), a framework, a goal, actions to reach this goal, and obstacles. Participants can be asked to evaluate the extent to which they agree with the choice of the persona or what they would do if they were in the situation of the persona. This distancing—giving advice about someone else's choice rather than answering for yourself—can help address sensitive themes [23,32].

Experience-Based Co-design

Experience-based co-design is a technique using narratives, usually video recorded, that allows patients and researchers to work in partnership to co-design new services and technology

with the objective of improving the quality of health services. This technique has been applied with success for fostering PPI [33-36].

Digital Methods to Promote PPI

Digital methods can be enablers of PPI, and the public can help modulate and develop digital technology [127]. Dedicated websites can enhance PPI not only by providing information, but also by organizing PPI itself. An example is the use of a webpage to organize a patient-led research hub. This is an initiative aimed at making patients and the public leaders of their own research projects. The researchers assist and support the patients and the public who proposed projects [37].

Crowdsourcing is a method by which many people are engaged on the web for a common goal such as obtaining new ideas and analyzing data. The main advantage is that it is possible to have a large number of contributors in a very short time [128]. However, because crowdsourcing participants are generally younger than participants from traditional nonweb-based procedures, these participants are not always representative of the target population [129].

Social media analysis is the use of digital data from social networks for epidemiological purposes. This information increases knowledge about epidemiological trends and may be very different from information obtained by traditional methods [130]. Social network data sources can be used, for example, to obtain information about the patient's research priorities [38]. This often generates large quantities of data and can offer the opportunity to use special techniques such as natural language processing for the analysis [131].

Web-based platforms are internet services where the public and innovators meet and can be created for allowing coworking [132]. Vasilica et al [39] reported creating a web-based network to co-design a social media–based platform with the aim of generating information to improve disease outcomes [39]. In addition, web-based platforms through a web-based voting system have been used for establishing research priorities [133].

Fields of Involvement

We found that many studies (60/97, 62%) involved patients and the public for generation of new ideas. Patients or the public contributed also in coproduction, co-design, or study scope (42/97, 43%) and in establishing research priorities (31/97, 32%). Other forms of PPI included participation in data analysis (25/97, 26%), as coauthors of a scientific article (17/97, 18%), as members of a steering committee or advisory group (16/97, 16%), reviewing or writing protocols (14/97, 14%), in the interpretation of results (12/97, 12%), in the dissemination of results and advocacy (11/97, 11%), in the data collection (9/97, 9%), in the development of the recruitment strategy (7/97, 7%), as a project manager (4/97, 4%), and being coinvestigator or having patient- or public-led projects (2/97, 2%). We found only 1 study where patients were involved in the co-design of mobile health tools (1/97, 1%). Finally, recent studies reported an early PPI in at least three different stages or fields of involvement and throughout the life of the research project (Table 1 and Figure 1).

Table 1. Fields of involvement in the included studies (N=97).

Type of involvement	Frequency, n (%)	References
Generation of new ideas	60 (62)	[18,20,22-27,30,32-36,38-82,134]
Coproduction, co-design, or study scope	42 (43)	[17,19,23,30,32-37,39,41,49,54,55,59,60,65-69,74,75,78,81,83-98]
Establishing research priorities	31 (32)	[24,26,29,30,34,35,41-43,46-48,53,54,57,60,62-64,66,72,75,77,81,83,89,90,99-102]
Participation in data analysis	25 (26)	[24,30,31,34,51,60,61,66,69-71,78,81,87,90,91,94,96,103-109]
As coauthors of a scientific article	17 (18)	[24,30,31,57,60,61,68,69,73,83,86,94-96,105,110,111]
As members of a steering committee or advisory group	16 (16)	[26,28,30,31,35,42,43,53,62,76,78,84,91,106,110,112]
Reviewing or writing protocols	14 (14)	[30,49,51,59,60,68,73,74,78,83,91,95,111,113]
Interpretation of results	12 (12)	[24,30,41,49,60,66,69,78,91,94,96,104]
Dissemination of results and advocacy	11 (11)	[30,39,41,49,57,60,70,78,86,94,104]
Data collection	9 (9)	[54,60,66,70,88,91,94,96,114]
Development of the recruitment strategy	7 (7)	[41,60,66,68,91,104,111]
Project management	4 (4)	[30,37,78,94]
Coinvestigator or having patient- or public-led projects	2 (2)	[37,94]
Co-design of mobile health tools	1 (1)	[135]
At least three different stages or fields of involvement and throughout the life of the research project	39 (40)	[24,26,28,30,31,34,35,39,41-43,49,51,53,54,57,59-61,66,68-70,73-75,78,81,83,86,89-91,94-96,104,110,111]

To develop a successful PPI project, patients or the public and researchers must have, or develop, certain skills. For example, researchers need to become familiar with PPI as a research approach, know how to manage a PPI project, and how to deal with conflict. As for the patients and the public, they must understand the research process and develop capacities for

management and conflict management. However, it is not mandatory that patients have specific vocational or educational training [136].

PPI in Epidemiology: Trials and Cohort Studies

Web-based trials are more and more frequently described in the literature. Price et al [137] performed a systematic review of web-based trials and found that PPI was only reported in 24% (10/41) of the trials included in the review. Face-to-face meetings and email contact were the most common ways of interaction [137].

Taylor et al [97] performed a cohort study in patients with cancer that involved the patients in the creation and choice of a brand for the cohort. With a 1-day workshop, patients and researchers co-designed the brand. The results showed higher acceptance and retention of the study than expected. An ongoing cohort study used social media (Facebook) for PPI by creating a closed group of patients and families to bring new ideas to the project [80]. Meetings with a family advisory committee were organized regularly.

Morris et al [116], in the context of an epidemiological study, investigated PPI with surveys and postevent interviews and wrote recommendations. Before an event, they suggest sending a detailed document with the topics to be discussed. During the event, they suggest having enough space between tables to allow all participants to be heard, providing materials to facilitate note taking, taking a whole day for the meeting, and arranging a facilitator for each table. Finally, after the event, they recommend a follow-up by sending the notes to the participants [116].

Barriers to PPI

Domecq et al [138] in a systematic review described some barriers to PPI. They highlighted 2 barriers: the excessive time taken for training activities and attendance and the risk of a tokenistic involvement. Another barrier reported in participants who were frail was frustration because of discontinuity in the involvement [139]. Barriers reported by researchers were concerns about the quality of research, ethical issues, lack of funding, failure of the PPI in the past, and not being convinced of the real need for PPI in the cohort [140]. Maccarthy et al [113] described communication issues as a key barrier to PPI. Researchers fear not being able to explain the project and not being able to engage patients and the public in the project; they also feel discomfort speaking about their experiences with patients and the public and fear having misunderstandings.

Facilitators of PPI

Creating a safe and welcoming environment where each contributor feels empowered and confident facilitates PPI [94]. The coproduction process can give participants the self-confidence to take responsibility for the entire duration of a project [78]. In addition, an iterative process of PPI evaluation during the entire research cycle has been proposed to ensure success in PPI [113]. Mathie et al [141] reported that feedback for patients, when provided, motivated them to continue their collaboration with researchers. Concerns for well-being, trust, mutual respect, and flexibility in time and methods were facilitators of PPI [139]. Chambers et al [142] found key areas that may facilitate or hinder the development of PPI. These key areas were the following: good role definition, recognition of difficulties, integration through organizations, training, developing networking, considering different perspectives, improving communication, and recognizing the relevance of emotional impact. Finally, concerning digital interventions and PPI, O'Connor et al [143] recommend investing in raising awareness of the usefulness of digital tools, improving health literacy, and using optimal tool design.

A Patient Advocate's Viewpoint of PPI in Diabetes Research

When writing this review, it was natural to allow a potential representative of study participants to express how they see PPI so far in research and what they are expecting from researchers to increase the participation of people with diabetes in research (Textbox 1, written by RS).

Textbox 1. Involving people with diabetes in research: a patient advocate's viewpoint.

Viewpoint of a patient advocate

- On November 14 each year, World Diabetes Day is celebrated across the globe. The International Diabetes Federation with the World Health Organization created this awareness campaign in 1991 to respond to the growing numbers of people with diabetes worldwide. In 2006, World Diabetes Day was deemed an official United Nations Day with a special resolution, becoming only the second health condition so recognized.
- Why November 14? That date marks the birthday of Sir Frederick Banting, who, along with Charles Best, is credited with the discovery of insulin. The day is acknowledged by diabetes organizations, health care professionals, researchers, and governments. Most importantly, people with diabetes have embraced the day to acknowledge, commemorate, and also celebrate life with diabetes as we gratefully signpost the man whose research is responsible for our very lives.
- People with diabetes are interested in research. We know that the developments we see each and every year that advance how we live with diabetes are the result of research. We are interested in the different branches of research—clinical, educational, social, and behavioral—because we know better than anyone that living with diabetes is a multi-pronged existence that affects every part of our lives.
- However, despite how much we appreciate the work of researchers and how keen we are to learn more, sometimes it is difficult to engage with us and involve us as participants. Let us explore how we can address this gap and consider some changes that can be made to encourage people with diabetes to take more interest in research.

Tell the story

- The story of Banting and Best is folklore for those of us living with diabetes. It is a compelling story, but so are many other research tales. Unfortunately, the narrative is not always told especially effectively. It is difficult to make research sound relevant to people living each day with diabetes when research involves cells in a petri dish or stem cells in a temperature-controlled laboratory. What is the difference this work will make to our day-to-day lives?
- It is exhausting for us to hear how mice are cured of diabetes (once again), especially when we know that our cure is still as elusive as ever.
- However, these stories—the cells and the mice—are links in a long chain that lead to significant developments that do directly affect us. At the moment, explaining that seems lost in translation, and researchers need to think about how to decode in basic language how the work they are doing has the potential to make significant changes to the everyday life of people with diabetes and that participating in relevant research gets us to that goal.
- Even those researchers whose work is more practical based are not always especially successful in describing the impact of their work on those of us with diabetes. Plain language statements are a start, but looking for even more nuanced and targeted ways to communicate is important.

Tell it in a tweet!

- With 280 characters on offer, Twitter is the perfect platform for researchers to hone their short story–telling skills. Practice the elevator pitch of your research by narrowing down the key points and benefit to people with diabetes, and share it on the web to encourage interest. (Twitter threads allow for linking a number of tweets together, so if you need more than 280 characters, you can take a couple of tweets. But do keep it brief!)

What is involved?

- When recruiting people for your research, be very clear about what they will need to do. How much time is involved? Where will they need to go? Will there be any invasive procedures and how uncomfortable are they likely to be? (Be honest!)
- Follow-up is critically important. A complaint we hear from people participating in research is that once their involvement is over, they never again hear from the research team. This can be especially frustrating if people have invested a lot of time and energy in a trial. Regular updates through a newsletter or social media page keep people informed and linked in with your work. This is especially important if you are planning to recruit people for future phases of your study.

Participants, not subjects #LanguageMatters

- The words you use when communicating to, and about, people with diabetes are critically important. Refer to language position statements developed by diabetes organizations to ensure that your language is supportive, empowering, positive, and encouraging. We people with diabetes are more inclined to be involved if we see a study that treats us with respect.

The 2 camps—and how to bridge them

- There seem to be 2 main camps when it comes to diabetes research. Some believe that the focus should be primarily on finding a cure for diabetes. This seems to be especially prevalent in the type 1 diabetes space, with much of this thinking led by parents of children living with diabetes. In 1970, it was these parents who founded the leading diabetes research organization, the Juvenile Diabetes Foundation (renamed the Juvenile Diabetes Research Foundation), now known as JDRF. The original organization's mission was very clear: to find a cure for diabetes. In recent years, however, the research funded by JDRF has branched out to include studies looking at improving management through technology and drugs.
- However, it is important to acknowledge the importance of research that looks at better management. Without this research, there would be no treatment for diabetes-related complications and we would not have technology such as home blood glucose meters, insulin pumps, continuous and flash glucose monitors, algorithms to automate insulin delivery...and we would still be using the same insulin from dogs that Banting and Best had used.

- Just as important is the growing body of work and researchers dedicated to researching the social, psychosocial, behavioral, and emotional aspects of living with diabetes. As anyone living with diabetes will tell you, this condition is never just about metrics. It is very much about our *headspace* and how we feel about living with diabetes.

Involve us

- When is the best time to start to involve people with diabetes in your research? It is probably already too late! Have you consulted us when you were establishing your study design? And back up a little more...is the research really something that is going to be of interest or benefit to people with diabetes. Is the problem you are looking to solve really a problem for us?

Patient advisory committees

- Many research bodies now require patient advisory committees to be established as part of the overall study. Done well, these groups can provide invaluable input for projects. Done badly, they are nothing more than a tick-the-box exercise. Ensure that there is funding available for travel, accommodation, and other expenses. Be clear about what you expect the patient advisory committee members to do and which aspects of the project they will be involved in. Remember that patient advisory committee members will most likely be volunteering their time. Their expertise and time should be reimbursed by honoraria or hourly payments.

Not enough money in the pot

- Research dollars are never enough, and each year, there are more researchers contending for elusive grants. When the results from grant rounds are shared, it seems that diabetes is repeatedly the *poor cousin* of health conditions, regularly being awarded significantly less money (with fewer successful grants) than conditions such as cancer and cardiovascular disease. In recent grant announcements from the National Health and Medical Research Council in Australia, only 16 diabetes grants worth Aus $13.5 million (US $9.6 million) were awarded compared with 69 grants worth Aus $52.9 million (US $37.8 million) awarded for cancer research.
- People with diabetes can help to advance the cause of diabetes grant applications by telling their stories. Perhaps one of the reasons that diabetes receives comparatively little of the research bucket is because we have not been all that successful in telling our stories. Instead, we have created a false image of diabetes as a hugely preventable, self-inflicted condition, resulting in government and other research bodies considering diabetes a less worthy condition to fund.
- Researchers are encouraged to work closely with people with diabetes to help tell the story of why their own research is important and how it has the potential to help in the lives of people affected by, or at risk for, diabetes. Humanizing the story is important—all too often, diabetes is presented in the media as a headless overweight body, which only adds to the stigma and image problem of the condition.

The story of hope

- Research is selling an important feature: hope. People with diabetes trade on hope; we look for it in research because we know that is what holds the key to improving outcomes, reducing burden, and making our diabetes lives easier. We want to be part of those discoveries that promise a better life, and we want to be involved in your research that will help us get there.

Illustration of PPI in a Digital Cohort Study

Digital cohort studies are longitudinal studies in which the data come either totally (e-cohort) or partially (hybrid: e-cohort and traditional cohort) from digital sources. Modern cohort studies increasingly incorporate digital tools such as data generated on the web and connected devices that allow much wider use of data generated for multiple projects [9].

We elaborated a strategy of PPI for a digital cohort study. Tables 2 and 3 show recommendations for PPI at all stages of research. This participation was defined in 2 categories: *Recommended participation activities*, in which patients and the public participate more passively, helping to generate research ideas and prioritize those ideas by participating in surveys, and *Recommended involvement activities*, in which patients actively participate in collaborative work with researchers on an equal footing. Examples of these activities are events, meetings, workshops, and focus groups. The chosen strategies for the digital cohort study are detailed in 32 actions in total for participation as well as participation and involvement, corresponding to the different steps of the research process, and are based on the current recommendations of Involve [7], our review results, and the viewpoint of a patient advocate (Textbox 1). In addition, for the realization of certain activities, we suggest a recommended duration of the activity based on our review of the literature and the point of view of the patient advocate.

Table 2. Recommendations for the promotion of patient and public involvement (PPI) projects: concrete examples for a digital cohort study. Steps to be taken before starting the cohort study.

Stages (Involve list)	Suggested actions	
	Recommended participation activities	Recommended involvement activities
Identifying and prioritizing research axes	• Web-based survey through social media: identification of research questions. Duration of the activity: 15 minutes	• Videoconference meeting: establishing an international scientific steering committee with researchers and patients as members. Duration of the activity: 1 hour. Preparation: read agenda that should be sent 1 day before the meeting • Meeting: identification of, and invitation to, a group of patients interested to be involved as patient partners (eg, through patient associations). Duration of the activity: 3 hours. Preparation: not needed. Venue: local patient association • Videoconference meeting and use of web-based collaboration tools: cowriting PPI plan for the cohort and submission to an ethics committee. Duration of the activity: 2 hours. Preparation: read the proposal draft sent and written by researchers 1 week before
Designing	• Web-based survey through smartphone app: identification of research questions, web-based survey with open-ended questions. Duration of the activity: 10 minutes • Web-based survey: ranking research questions (through smartphone app, web-based survey using persona-scenario technique) with closed-ended questions. Duration of the activity: 15 minutes	• Web-based or mobile workshops: coproduction by giving feedback on study design and chosen questionnaires and research tools (such as mock-ups of app, user experience and user interface). Duration of the activity: 3 hours • Web-based training: language matters. Searching, choosing, and checking the most appropriate use of language for communication with the public and patients. Duration of the activity: 2 hours • Mobile focus group and survey (Multimedia Appendices 2 and 3): assessment of the beta version of smartphone app and flyers, as well as assessment of the wording and visual of the website, flyer, study objectives, and PPI expectations. Duration of the activity: 4 hours. Meeting place: comfortable, with catering and parking lot available
Drafting grant protocol	—[a]	• Web-based meeting: cowriting study protocol, involvement of patient associations as partners in grants. Duration of the activity: 1 hour. Preparation: read agenda that should be sent 1 day before the meeting
Testing and scaling up	• Surveys through smartphone app: testing of pilot study by limited number of potential study participants. Duration of the activity: 30 minutes	• Web-based meeting: co-designing pilot study on smartphone app. Duration of the activity: 1 hour. Preparation: read agenda that should be sent 1 day before the meeting • Web-based meeting: co-design of generalization phase and recruitment. Duration of the activity: 1 hour. Preparation: read agenda that should be sent 1 day before the meeting • Web-based meeting: advertise through social media and press for patients and the public to participate in the study

[a]No specific recommendations.

Table 3. Recommendations for the promotion of patient and public involvement (PPI) projects: concrete examples for a digital cohort study. Steps to be taken during the cohort study.

Stages (Involve list)	Suggested actions	
	Recommended participation activities	Recommended involvement activities
Analyzing and interpreting	—[a]	• Web-based meeting using web-based collaboration tools: cowriting of annual reports • Webpage, web-based workshops, meetings, and web-based collaboration tools: coproduction of research projects through a patient-led research hub. Webpage with a dedicated section for submission of projects by patient. Projects assessed by the scientific steering committee • Web-based workshops, meetings, and web-based collaboration tools: data analysis and interpretation of results • Meetings and web-based collaboration tools: writing of manuscripts cowritten by scientists and patients
Disseminating	—	• Social media: dissemination of publications coauthored by scientists and patients • Web-based meetings, workshops, and web-based collaboration tools: participation at conferences as author or coauthor • Focus groups and workshops: communication of research results (plain language, infographic, and dissemination)
Implementing	—	• Web-based meetings and workshops: implementation of some results from the study at hospitals and consultations facilitated by patients
Monitoring and evaluating	• Smartphone app: improving participants' retention by reminders • Email newsletter and social media announcement: follow-up of the project by researchers (once a month) • One-day general public event: follow-up of the project by researchers (once a year) • Web-based survey through social media and smartphone app: monitoring evolution of the research protocol (adding or deleting research questions)	• Web-based meetings: checking of data collection and data quality by patients • Web-based meetings and web-based collaboration tools: review of research projects by patient representatives • Social media: improving participants' retention by involved patients • One-day general public event: follow-up of the project by patients and public (once a year). Remuneration or facilities for attending should be budgeted for members of the scientific committee • Workshop: evaluation of PPI by researcher and contributors (use GRIPP2[b] checklist). Duration of the activity: 1 day. Meeting place: comfortable with catering and available parking places • Meetings and workshops: Monitoring evolution of the research protocol (adding or deleting research questions) by scientific steering committee (scientists and patient members)

[a]No specific recommendations.

[b]GRIPP2: revised version of Guidance for Reporting Involvement of Patients and the Public.

The advantage of a digital cohort study is that digital tools can be used to promote PPI at each stage as a primary method or as a complementary or alternative method. A digital cohort can facilitate PPI, allowing participation from remote locations, using a smartphone app with web-based questionnaires, organizing most of the meetings through videoconferences, and using web-based tools for coworking. However, we think that face-to-face activities are also recommended and these 2 approaches may be complementary.

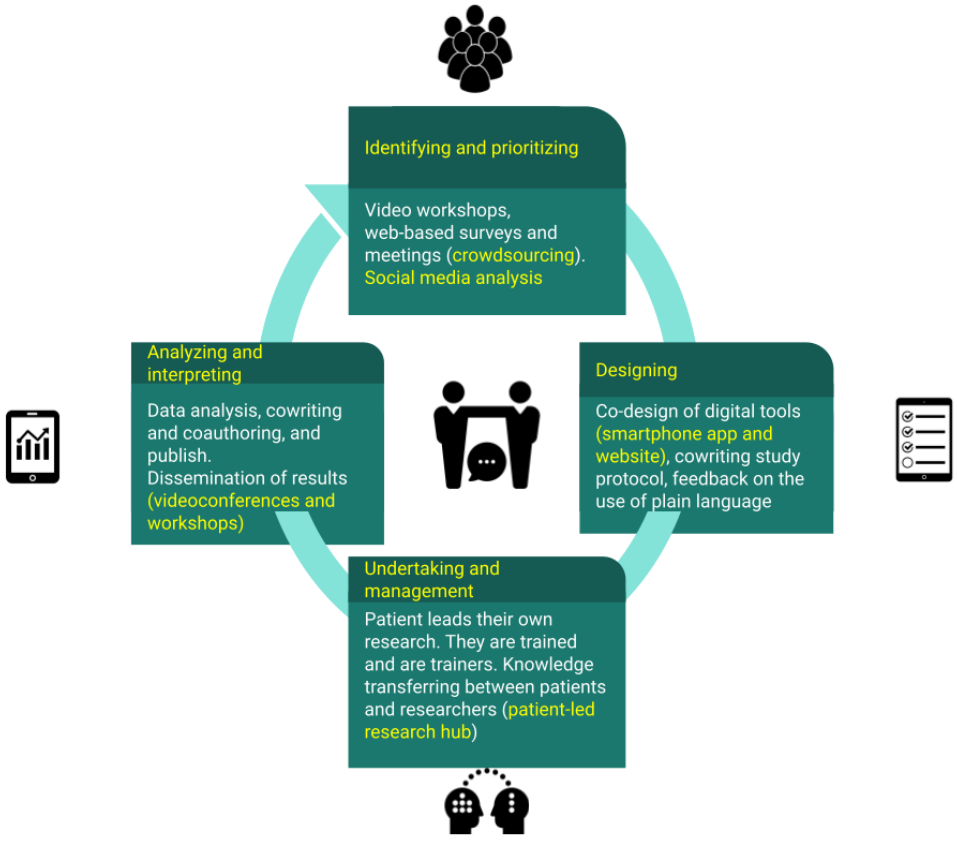
The recommendations are as follows:

- Recommendation 1: Identify patients and the public who might be interested in participating as members of the patient advisory steering committee. This contact can be achieved through social media and patients' organizations. Use digital tools to identify people. For example, contact organizations that are active on forums or social media.
- Recommendation 2: Write a PPI protocol in the digital cohort protocol describing all planned activities and include funding for patients. The patients who are part of the steering committee should be actively involved in this activity.
- Recommendation 3: Identify patients and the public who might be interested in participating in focus groups, semistructured interviews, or workshops by being involved actively in the design of the app, website, research ideas, and project evaluation.
- Recommendation 4: Organize focus groups as an important activity to obtain information about how PPI can be integrated in a digital cohort study. We present an example guide for this activity in Multimedia Appendix 2. A passionate and enthusiast moderator is needed. Sometimes it is not possible to find a place and a time that works for everyone. In such cases, mobile or web-based focus groups can be organized.

- Recommendation 5: Involve patients in the design of the smartphone app or other digital tools. For working on the design of a smartphone app, we encourage the use of an evaluation grid to assess the app with direct observation of how patients and the public use it. We present an example of an evaluation grid in Multimedia Appendix 3. Patients and the public can also assess a smartphone app using a validated scale. We suggest the application of the user version of the Mobile Application Rating Scale [144].
- Recommendation 6: Involve patients in the generation of new research ideas and research priorities. Using a smartphone app or through a website, we recommend that during the study, participants be invited to propose research questions at any time and as often as they wish. The patient advisory steering committee is involved in all the discussions about research priorities and follow-up of the project.
- Recommendation 7: Give feedback to patients and the public through web-based newsletters and social media. Evaluate regularly satisfaction with PPI. To assess the impact of PPI among active participants and to define roles and changes in the project, we advise that annual workshops should be organized. At these annual workshops, PPI can also be assessed by contributors (researchers, patients, and the public) using the GRIPP2 checklist [16].
- Recommendation 8: Establish a patient-researcher partnership. On the cohort's website and in the smartphone app, we encourage the creation of a *patient- or public-led research hub* space. This space can receive applications from patients or the public to develop research projects. The scientific steering committee assesses applications concerning scientific content and feasibility. The research team helps patients and the public to carry out their own projects by providing methodological support.
- Recommendation 9: Finally, we recommend inviting patient advocates, patients, or caregivers to become *study ambassadors*. They are volunteers actively involved in the study and social networks. The role of the study ambassadors will be to actively disseminate the results of the study and invite their networks to join the project.

Figure 2 shows our vision of integrating PPI in a digital cohort study in the whole research cycle.

Figure 2. Patient and public involvement in the research cycle of a digital cohort study. Digital tools are integrated at each stage of the research cycle, and some examples of digital tools are shown in the figure.



Discussion

Principal Findings

We observed that the most popular methods for PPI are workshops, focus groups, interviews, and surveys. The appropriate method must be carefully chosen to fit the research objective. For example, workshops and focus groups can be a rich source of information in the prestudy phase; they can be adapted to the context of the participants; and they allow direct observation of, and interaction among, participants. The downside is that they are highly dependent on the capacity of the moderators.

We have also described digital tools to collect information from the patients and the public for research, such as social media

listening and a particular web-based crowdsourcing survey with the *magic wand*–like question [24]. We think that digital tools may not only facilitate patient engagement in research, but can also stimulate a continuous and long-term participation and involvement of patients and the public.

Overall, none of these methods can address all relevant questions for scientists and patients or the public related to a research project. Therefore, we recommend a mix of methods to obtain optimal and meaningful information for the research project at stake. Different methods can be applied at different times of the project depending on the objective. Regarding the timing of introducing PPI, we recommend that PPI be included from the beginning (eg, co-design and research plan) to the end (publication and dissemination), followed by a PPI assessment.

We suggest pragmatic steps to integrate PPI in future epidemiological studies. These recommendations are as generic as possible but may not be applicable in all cases. They are based on current guidelines for PPI with concrete examples for a digital cohort study [14,15]. These recommendations should be understood as modular, meaning that they must be adapted to the study design, population, available budget, duration of the project, and local context. For PPI report and assessment, we recommend using the GRIPP2 checklist, a dedicated PPI reporting checklist [16].

Comparison With Prior Work

Domecq et al [138], in a systematic review, identified barriers to, and facilitators of, PPI. They found no evidence indicating which method was best for PPI. In our review, we found that a combination of methods is more common with greater involvement (patients included in many areas or across the research cycle). We integrated the barriers and facilitators described by Domecq et al [138] in our recommendations.

Liabo et al [145] performed a systematic review of good PPI practices. They compared the results of the systematic review with 3 involvement groups and found that the priorities were similar. However, the involvement groups found additional values that were not described in the literature, such as the enthusiasm of the participants and the choice of welcoming venues for the meetings. We integrated these reflections in our recommendations.

Nunn et al [146] reviewed reports of 96 human genomic research projects and found that only 33% (32/96) declared PPI. From these, most of the PPI activities were organized in formal groups (20/32, 63%), with 22% (7/32) using web-based tools (website, social media, and web-based communities). We found similar results with social media reporting in 6% (6/97) of the studies with PPI. We think that there is room for more PPI using digital tools.

Miah et al [147] conducted a scoping review of PPI in dementia research and found 19 studies from the United Kingdom and 1 from the Netherlands. Biddle et al [13] found an uneven distribution of PPI in Europe. They attributed this to a lack of infrastructure, support, and guidance. An example of support is that research-funding institutions in the United Kingdom require PPI in project applications. However, funding agencies in many countries do not have this requirement. We found similar results with most of the studies on PPI from the United Kingdom. We believe that PPI is still underreported or not performed in many countries.

Few epidemiological or clinical studies report PPI in the research process [11,146]. Studies that include patients and the public most often involve them only at the stage of idea generation, but not in the whole research cycle. This suggests that PPI is symbolically added or very limited.

Individual interviews are useful for tackling sensitive questions because participants do not feel dominated or influenced by the opinion of other participants. Surveys (web-based, telephone-based, or paper-based) are a pragmatic method of obtaining large-scale information from many people quickly. Differently, a survey can be very useful for generating new ideas and for refining and improving them when a new survey containing the generated ideas is launched for another sample of people.

PPI is a *project in the project*, creating its own challenges such as completing appropriate regulatory tasks and obtaining approval from an ethics committee. In the informed consent, the nature of participation must be specified, with clear and fair terms and conditions. This includes whether remuneration, cost coverage for the patient organization, and travel costs of patients or the public are provided and whether there are other benefits for the patients, such as nonfinancial compensation for the time allocated to project participation [14].

Several reasons for a lack of, or delay in, PPI have been mentioned in the literature. Some researchers do not include patients or the public in the research, arguing that the patient or public point of view is too subjective [148]. Furthermore, researchers may consider PPI only as a requirement of funding agencies for the project to be approved; therefore, PPI arrives late in the process and is treated as an afterthought [149]. We believe that obstacles can be overcome when they are identified and taken into account in the plan and recommendations.

When PPI is considered, there is the risk of selective PPI, which means that only those within the community who agreed with the research objectives are included. In addition, this may have the risk of hearing only the opinion of the most active patients or public, which may not reflect the opinion of most of the other patients and can therefore lead to research designs that still do not research the questions that interest most patients. We recommend nonselective PPI, meaning that the aim should be to involve a diverse selection of participants in the PPI process, including patient organizations, patient advocates who are legitimately speaking on behalf of a patient community, and individual patients, to ensure that the opinions and views of the participants are representative of most patients with the same condition [47].

Researchers should make a greater effort to minimize the burden of participation for patients and maximize the benefits for participants at the same time. A lack of participation is more likely to occur when there is poor or 1-way communication, resulting in poorly organized protocols and demanding follow-ups that lead to noninformed, noninvolved, and nonmotivated participants. Nevertheless, when done properly,

patients can associate their participation with feelings such as usefulness, empowerment, and consideration [150].

Limitations

This study includes certain limitations. Our search was limited to studies in English. In addition, a part of what is happening in PPI is described in the gray literature and has therefore been excluded from the review. Our review may give the impression that early involvement in setting research priorities is the norm, but this may not be true because of a potential publication bias. For example, the results of studies with PPI are often not intended to show improvements in efficacy in clinical studies; therefore, they are less likely to be accepted by publishers for publication or publication may be delayed because of negative or statistically insignificant results.

Conclusions

There are, and rightly so, many expectations on the part of patients and the public to be actively involved in research and not only by providing data, but also as research partners. PPI can contribute to patient empowerment by increasing disease awareness and according recognition as actors of their own condition [151]. However, PPI is uneven among countries and research institutions, and even now many patients and the public are not yet involved in research [13] and ignore or do not have access to research protocols [152]. Digital tools such as websites, social media, and connected devices have been increasingly incorporated into cohort studies and could be leveraged to increase PPI [153]. Digital tools can facilitate PPI by providing an opportunity for remote access and therefore easier participation. In addition, digital tools can facilitate PPI by enabling feedback and interaction between researchers and patient collaborators [154].

PPI can be a powerful approach to increase the relevance of research projects. We have shown that PPI must be planned in the initial phases of the development of a new epidemiological study and then be considered throughout the life of the research project. Combining different approaches of PPI seems to be the most effective strategy for improving the quality of research.

Some techniques such as *persona-scenario* are very powerful for idea generation and can be combined with digital tools. In addition, web-based surveys are easy to implement and allow involving many participants (crowdsourcing).

Digital methods such as social media listening or web-based *magic wand*–like questions can also offer useful complementary channels of interaction and help to identify key information such as research gaps at large scale with a limited cost directly from the target population. As such, we recommend that these methods be also integrated in the PPI process. With the example of a new digital cohort study, we offer practical guidelines to implement and run a patient- or public-centered research study. We therefore encourage investigators to rely on our practical recommendations to increase PPI in future epidemiological studies.

Acknowledgments

This study was funded by the Ministry of Higher Education and Research of Luxembourg and the Luxembourg Institute of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors thank Ms Sandrine Lavalle for her support for this work.

Authors' Contributions

GAA reviewed the literature, selected the studies, extracted and analyzed data, and wrote the first draft. RS wrote Textbox 1 and contributed to the review and editing of the manuscript. CG, AF, TS, VTT, PR, TB, and LH contributed to the writing and editing of the manuscript. GF designed the study, selected and analyzed the data, and contributed to the writing and editing of the manuscript. All authors agreed on the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methods and fields of patient and public involvement.
[DOCX File , 152 KB - jmir_v23i12e25743_app1.docx]

Multimedia Appendix 2
A focus group guide.
[DOCX File , 13 KB - jmir_v23i12e25743_app2.docx]

Multimedia Appendix 3
Grid for assessment of a smartphone app.
[DOCX File , 15 KB - jmir_v23i12e25743_app3.docx]

References

1. Tips sheet: recruiting members of the public to get involved in research funding and commissioning processes. INVOLVE - National Institute for Health Research. 2012. URL: http://www.invo.org.uk/wp-content/uploads/2012/04/Recruitment-tips-sheet.pdf [accessed 2019-11-20]
2. What is patient and public involvement and public engagement? National Institute for Health Research. 2020. URL: https://www.spcr.nihr.ac.uk/PPI/what-is-patient-and-public-involvement-and-engagement [accessed 2021-12-07]
3. Trish Greenhalgh: towards an institute for patient-led research. The BMJ Opinion. 2019. URL: https://blogs.bmj.com/bmj/2019/11/12/trisha-greenhalgh-towards-an-institute-for-patient-led-research/ [accessed 2021-12-07]
4. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet 2014 Jan 11;383(9912):156-165. [doi: 10.1016/S0140-6736(13)62229-1] [Medline: 24411644]
5. Crocker JC, Ricci-Cabello I, Parker A, Hirst JA, Chant A, Petit-Zeman S, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. Br Med J 2018 Nov 28;363:k4738 [FREE Full text] [doi: 10.1136/bmj.k4738] [Medline: 30487232]
6. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
7. Hayes H, Buckland S, Tarpey M. Briefing Notes for Researchers: Public Involvement in NHS, Public Health and Social Care Research. Eastleigh: INVOLVE; 2012.
8. 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]
9. Fagherazzi G, Ravaud P. Digital diabetes: perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
10. Boote J, Baird W, Beecroft C. Public involvement at the design stage of primary health research: a narrative review of case examples. Health Policy 2010 Apr;95(1):10-23. [doi: 10.1016/j.healthpol.2009.11.007] [Medline: 19963299]
11. Harris J, Haltbakk J, Dunning T, Austrheim G, Kirkevold M, Johnson M, et al. How patient and community involvement in diabetes research influences health outcomes: a realist review. Health Expect 2019 Oct;22(5):907-920 [FREE Full text] [doi: 10.1111/hex.12935] [Medline: 31286639]
12. Price A, Schroter S, Snow R, Hicks M, Harmston R, Staniszewska S, et al. Frequency of reporting on patient and public involvement (PPI) in research studies published in a general medical journal: a descriptive study. BMJ Open 2018 Dec 23;8(3):e020452 [FREE Full text] [doi: 10.1136/bmjopen-2017-020452] [Medline: 29572398]
13. Biddle MS, Gibson A, Evans D. Attitudes and approaches to patient and public involvement across Europe: a systematic review. Health Soc Care Community 2021 Jan;29(1):18-27. [doi: 10.1111/hsc.13111] [Medline: 32705752]
14. Hunter A, Facey K, Thomas V, Haerry D, Warner K, Klingmann I, et al. EUPATI guidance for patient involvement in medicines research and development: health technology assessment. Front Med (Lausanne) 2018 Sep 6;5:231 [FREE Full text] [doi: 10.3389/fmed.2018.00231] [Medline: 30238004]
15. INVOLVE. Briefing Notes for Researchers: Public Involvement in NHS, Public Health and Social Care Research. Eastleigh: INVOLVE; 2012.
16. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. Res Involv Engagem 2017;3:13 [FREE Full text] [doi: 10.1186/s40900-017-0062-2] [Medline: 29062538]
17. Mackintosh N, Sandall J, Collison C, Carter W, Harris J. Employing the arts for knowledge production and translation: visualizing new possibilities for women speaking up about safety concerns in maternity. Health Expect 2018 Jun 17;21(3):647-658 [FREE Full text] [doi: 10.1111/hex.12660] [Medline: 29341347]
18. Kelemen M, Surman E, Dikomitis L. Cultural animation in health research: an innovative methodology for patient and public involvement and engagement. Health Expect 2018 Aug;21(4):805-813 [FREE Full text] [doi: 10.1111/hex.12677] [Medline: 29532582]
19. Evans BA, Porter A, Snooks H, Burholt V. A co-produced method to involve service users in research: the SUCCESS model. BMC Med Res Methodol 2019 Feb 15;19(1):34 [FREE Full text] [doi: 10.1186/s12874-019-0671-6] [Medline: 30770732]
20. Eccles A, Bryce C, Turk A, Atherton H. Patient and public involvement mobile workshops - convenient involvement for the un-usual suspects. Res Involv Engagem 2018;4:38 [FREE Full text] [doi: 10.1186/s40900-018-0123-1] [Medline: 30386631]
21. How do I hold a PPI meeting using virtual tools? National Institute for Health Research. 2020. URL: https://www.spcr.nihr.ac.uk/PPI/resources-for-researchers/faq/how-do-i-hold-ppi-meetings-using-virtual-tools [accessed 2020-10-22]
22. Boddy K, Cowan K, Gibson A, Britten N. Does funded research reflect the priorities of people living with type 1 diabetes? A secondary analysis of research questions. BMJ Open 2017 Sep 27;7(9):e016540 [FREE Full text] [doi: 10.1136/bmjopen-2017-016540] [Medline: 28963289]

23. Knowles S, Hays R, Senra H, Bower P, Locock L, Protheroe J, et al. Empowering people to help speak up about safety in primary care: using codesign to involve patients and professionals in developing new interventions for patients with multimorbidity. Health Expect 2018 Apr;21(2):539-548 [FREE Full text] [doi: 10.1111/hex.12648] [Medline: 29266797]
24. Tran V, Riveros C, Péan C, Czarnobroda A, Ravaud P. Patients' perspective on how to improve the care of people with chronic conditions in France: a citizen science study within the ComPaRe e-cohort. BMJ Qual Saf 2019 Nov;28(11):875-886 [FREE Full text] [doi: 10.1136/bmjqs-2018-008593] [Medline: 31015376]
25. Little P, Everitt H, Williamson I, Warner G, Moore M, Gould C, et al. Preferences of patients for patient centred approach to consultation in primary care: observational study. Br Med J 2001 Feb 24;322(7284):468-472 [FREE Full text] [doi: 10.1136/bmj.322.7284.468] [Medline: 11222423]
26. Boney O, Bell M, Bell N, Conquest A, Cumbers M, Drake S, et al. Identifying research priorities in anaesthesia and perioperative care: final report of the joint National Institute of Academic Anaesthesia/James Lind Alliance Research Priority Setting Partnership. BMJ Open 2015 Dec 16;5(12):e010006 [FREE Full text] [doi: 10.1136/bmjopen-2015-010006] [Medline: 26674506]
27. Higgins PS, Shugrue N, Ruiz K, Robison J. Medicare and Medicaid users speak out about their health care: the real, the ideal, and how to get there. Popul Health Manag 2015 Apr;18(2):123-130. [doi: 10.1089/pop.2014.0056] [Medline: 25247347]
28. Howe A, Mathie E, Munday D, Cowe M, Goodman C, Keenan J, et al. Learning to work together - lessons from a reflective analysis of a research project on public involvement. Res Involv Engagem 2017;3:1 [FREE Full text] [doi: 10.1186/s40900-016-0051-x] [Medline: 29062526]
29. Rawson TM, Castro-Sánchez E, Charani E, Husson F, Moore LS, Holmes AH, et al. Involving citizens in priority setting for public health research: implementation in infection research. Health Expect 2018 Feb;21(1):222-229 [FREE Full text] [doi: 10.1111/hex.12604] [Medline: 28732138]
30. Dewa LH, Lawrence-Jones A, Crandell C, Jaques J, Pickles K, Lavelle M, et al. Reflections, impact and recommendations of a co-produced qualitative study with young people who have experience of mental health difficulties. Health Expect 2020 Jun 09. [doi: 10.1111/hex.13088] [Medline: 32515538]
31. O'Hara MC, Hynes L, O'Donnell M, Keighron C, Allen G, Caulfield A, with the D1 Now Type 1 Diabetes Young Adult Study Group. Strength in Numbers: an international consensus conference to develop a novel approach to care delivery for young adults with type 1 diabetes, the D1 Now Study. Res Involv Engagem 2017;3:25 [FREE Full text] [doi: 10.1186/s40900-017-0076-9] [Medline: 29214056]
32. Valaitis R, Longaphy J, Ploeg J, Agarwal G, Oliver D, Nair K, et al. Health TAPESTRY: co-designing interprofessional primary care programs for older adults using the persona-scenario method. BMC Fam Pract 2019 Sep 04;20(1):122 [FREE Full text] [doi: 10.1186/s12875-019-1013-9] [Medline: 31484493]
33. Locock L, Robert G, Boaz A, Vougioukalou S, Shuldham C, Fielden J, et al. Using a national archive of patient experience narratives to promote local patient-centered quality improvement: an ethnographic process evaluation of 'accelerated' experience-based co-design. J Health Serv Res Policy 2014 Oct 19;19(4):200-207. [doi: 10.1177/1355819614531565] [Medline: 24840387]
34. Hatton AL, Haslam C, Bell S, Langley J, Woolrych R, Cory C, et al. Innovative solutions to enhance safe and green environments for ageing well using co-design through patient and public involvement. Res Involv Engagem 2020;6:45 [FREE Full text] [doi: 10.1186/s40900-020-00223-4] [Medline: 32760595]
35. Hyatt A, Morkunas B, Davey D, Thai AA, Trewhella M, Duffy M, et al. Co-design and development of online video resources about immunotherapy with patients and their family. Patient Educ Couns 2020 Sep 16. [doi: 10.1016/j.pec.2020.09.014] [Medline: 32988686]
36. Mulvale G, Green J, Miatello A, Cassidy AE, Martens T. Finding harmony within dissonance: engaging patients, family/caregivers and service providers in research to fundamentally restructure relationships through integrative dynamics. Health Expect 2020 Jun 11. [doi: 10.1111/hex.13063] [Medline: 32529748]
37. Mader LB, Harris T, Kläger S, Wilkinson IB, Hiemstra TF. Inverting the patient involvement paradigm: defining patient led research. Res Involv Engagem 2018 Jul;4:21 [FREE Full text] [doi: 10.1186/s40900-018-0104-4] [Medline: 30002875]
38. Priest C, Knopf A, Groves D, Carpenter JS, Furrey C, Krishnan A, et al. Finding the patient's voice using big data: analysis of users' health-related concerns in the ChaCha question-and-answer service (2009-2012). J Med Internet Res 2016;18(3):e44 [FREE Full text] [doi: 10.2196/jmir.5033] [Medline: 26960745]
39. Vasilica CM, Brettle A, Ormandy P. A co-designed social media intervention to satisfy information needs and improve outcomes of patients with chronic kidney disease: longitudinal study. JMIR Form Res 2020 Jan 27;4(1):e13207 [FREE Full text] [doi: 10.2196/13207] [Medline: 32012040]
40. Piercy H, Yeo M, Yap S, Hart AR. What are the information needs of parents caring for a child with Glutaric aciduria type 1? BMC Pediatr 2019 Oct 13;19(1):349 [FREE Full text] [doi: 10.1186/s12887-019-1742-x] [Medline: 31607269]
41. Synnot A, Bragge P, Lowe D, Nunn JS, O'Sullivan M, Horvat L, et al. Research priorities in health communication and participation: international survey of consumers and other stakeholders. BMJ Open 2018 May 08;8(5):e019481 [FREE Full text] [doi: 10.1136/bmjopen-2017-019481] [Medline: 29739780]

42. Finer S, Robb P, Cowan K, Daly A, Shah K, Farmer A. Setting the top 10 research priorities to improve the health of people with Type 2 diabetes: a Diabetes UK-James Lind Alliance Priority Setting Partnership. Diabet Med 2018 Jul 15;35(7):862-870 [FREE Full text] [doi: 10.1111/dme.13613] [Medline: 29485717]
43. Bethell J, Puts MT, Sattar S, Andrew MK, Choate AS, Clarke B, et al. The Canadian frailty priority setting partnership: research priorities for older adults living with frailty. Can Geriatr J 2019 Mar;22(1):23-33 [FREE Full text] [doi: 10.5770/cgj.22.336] [Medline: 31501680]
44. Berwick DM, Loehrer S, Gunther-Murphy C. Breaking the rules for better care. J Am Med Assoc 2017 Jun 06;317(21):2161-2162. [doi: 10.1001/jama.2017.4703] [Medline: 28448652]
45. Zimbudzi E, Lo C, Robinson T, Ranasinha S, Teede HJ, Usherwood T, et al. The impact of an integrated diabetes and kidney service on patients, primary and specialist health professionals in Australia: a qualitative study. PLoS One 2019;14(7):e0219685 [FREE Full text] [doi: 10.1371/journal.pone.0219685] [Medline: 31306453]
46. Gregg A, Getz N, Benger J, Anderson A. A novel collaborative approach to building better clinical trials: new insights from a patient engagement workshop to propel patient-centricity forward. Ther Innov Regul Sci 2019 May 22:2168479019849875. [doi: 10.1177/2168479019849875] [Medline: 31117820]
47. Russell G, Starr S, Elphick C, Rodogno R, Singh I. Selective patient and public involvement: the promise and perils of pharmaceutical intervention for autism. Health Expect 2018 Apr;21(2):466-473 [FREE Full text] [doi: 10.1111/hex.12637] [Medline: 29090494]
48. Woodward L, Johnson S, Walle JV, Beck J, Gasteyger C, Licht C, aHUS Registry SAB. An innovative and collaborative partnership between patients with rare disease and industry-supported registries: the Global aHUS Registry. Orphanet J Rare Dis 2016 Nov 21;11(1):154 [FREE Full text] [doi: 10.1186/s13023-016-0537-5] [Medline: 27871301]
49. Westfall JM, VanVorst RF, Main DS, Herbert C. Community-based participatory research in practice-based research networks. Ann Fam Med 2006;4(1):8-14 [FREE Full text] [doi: 10.1370/afm.511] [Medline: 16449391]
50. Neville C, Da CD, Mill C, Rochon M, Aviña-Zubieta JA, Pineau CA, et al. The needs of persons with lupus and health care providers: a qualitative study aimed toward the development of the Lupus Interactive Navigator™. Lupus 2014 Feb;23(2):176-182. [doi: 10.1177/0961203313517154] [Medline: 24335012]
51. Jørgensen CR, Eskildsen NB, Johnsen AT. User involvement in a Danish project on the empowerment of cancer patients - experiences and early recommendations for further practice. Res Involv Engagem 2018;4:26 [FREE Full text] [doi: 10.1186/s40900-018-0105-3] [Medline: 30123531]
52. Casanova T, Black C, Rafiq S, Hugill-Jones J, Read JC, Vancleef K. The impact of active research involvement of young children in the design of a new stereotest. Res Involv Engagem 2020;6:29 [FREE Full text] [doi: 10.1186/s40900-020-00194-6] [Medline: 32518689]
53. Birnie KA, Dib K, Ouellette C, Dib MA, Nelson K, Pahtayken D, et al. Partnering for Pain: a Priority Setting Partnership to identify patient-oriented research priorities for pediatric chronic pain in Canada. CMAJ Open 2019;7(4):654-664 [FREE Full text] [doi: 10.9778/cmajo.20190060] [Medline: 31699686]
54. Brainard JS, Al Assaf E, Omasete J, Leach S, Hammer CC, Hunter PR. Forced migrants involved in setting the agenda and designing research to reduce impacts of complex emergencies: combining with patient and public involvement. Res Involv Engagem 2017;3:23 [FREE Full text] [doi: 10.1186/s40900-017-0073-z] [Medline: 29142759]
55. Brett L, Nguyen AD, Siette J, Dove-Pizarro J, Hourihan F, Georgiou A. The co-design of timely and meaningful information needed to enhance social participation in community aged care services: think tank proceedings. Australas J Ageing 2020 Mar;39(1):162-167. [doi: 10.1111/ajag.12706] [Medline: 31411384]
56. Brown K, Dyas J, Chahal P, Khalil Y, Riaz P, Cummings-Jones J. Discovering the research priorities of people with diabetes in a multicultural community: a focus group study. Br J Gen Pract 2006 Mar;56(524):206-213 [FREE Full text] [Medline: 16536961]
57. Costello W, Dorris E. Laying the groundwork: building relationships for public and patient involvement in pre-clinical paediatric research. Health Expect 2020 Feb;23(1):96-105 [FREE Full text] [doi: 10.1111/hex.12972] [Medline: 31625656]
58. Coultas C, Kieslich K, Littlejohns P. Patient and public involvement in priority-setting decisions in England's Transforming NHS: an interview study with Clinical Commissioning Groups in South London sustainability transformation partnerships. Health Expect 2019 Dec;22(6):1223-1230 [FREE Full text] [doi: 10.1111/hex.12948] [Medline: 31410967]
59. Coupe N, Mathieson A. Patient and public involvement in doctoral research: impact, resources and recommendations. Health Expect 2020 Feb;23(1):125-136 [FREE Full text] [doi: 10.1111/hex.12976] [Medline: 31613049]
60. Dawson S, Ruddock A, Parmar V, Morris R, Cheraghi-Sohi S, Giles S, et al. Patient and public involvement in doctoral research: reflections and experiences of the PPI contributors and researcher. Res Involv Engagem 2020;6:23 [FREE Full text] [doi: 10.1186/s40900-020-00201-w] [Medline: 32426162]
61. Frost J, Gibson A, Harris-Golesworthy F, Harris J, Britten N. Patient involvement in qualitative data analysis in a trial of a patient-centred intervention: reconciling lay knowledge and scientific method. Health Expect 2018 Dec;21(6):1111-1121 [FREE Full text] [doi: 10.1111/hex.12814] [Medline: 30073734]
62. Gadsby R, Snow R, Daly AC, Crowe S, Matyka K, Hall B, et al. Setting research priorities for Type 1 diabetes. Diabet Med 2012 Oct;29(10):1321-1326. [doi: 10.1111/j.1464-5491.2012.03755.x] [Medline: 22823450]

63. Ghisoni M, Wilson CA, Morgan K, Edwards B, Simon N, Langley E, et al. Priority setting in research: user led mental health research. Res Involv Engagem 2017;3:4 [FREE Full text] [doi: 10.1186/s40900-016-0054-7] [Medline: 29062529]
64. Gillard S, Simons L, Turner K, Lucock M, Edwards C. Patient and public involvement in the coproduction of knowledge: reflection on the analysis of qualitative data in a mental health study. Qual Health Res 2012 Aug;22(8):1126-1137. [doi: 10.1177/1049732312448541] [Medline: 22673090]
65. Grant C, Widnall E, Cross L, Simonoff E, Downs J. Informing the development of an E-platform for monitoring wellbeing in schools: involving young people in a co-design process. Res Involv Engagem 2020;6:51 [FREE Full text] [doi: 10.1186/s40900-020-00219-0] [Medline: 32908677]
66. Grundy A, Keetharuth AD, Barber R, Carlton J, Connell J, Buck ET, et al. Public involvement in health outcomes research: lessons learnt from the development of the recovering quality of life (ReQoL) measures. Health Qual Life Outcomes 2019 Apr 11;17(1):60 [FREE Full text] [doi: 10.1186/s12955-019-1123-z] [Medline: 30975153]
67. Irving A, Turner J, Marsh M, Broadway-Parkinson A, Fall D, Coster J, et al. A coproduced patient and public event: an approach to developing and prioritizing ambulance performance measures. Health Expect 2018 Feb;21(1):230-238 [FREE Full text] [doi: 10.1111/hex.12606] [Medline: 28841272]
68. Madden M, Morris S, Ogden M, Lewis D, Stewart D, McCambridge J. Producing co-production: reflections on the development of a complex intervention. Health Expect 2020 Jun;23(3):659-669 [FREE Full text] [doi: 10.1111/hex.13046] [Medline: 32233053]
69. Mann C, Chilcott S, Plumb K, Brooks E, Man M. Reporting and appraising the context, process and impact of PPI on contributors, researchers and the trial during a randomised controlled trial - the 3D study. Res Involv Engagem 2018;4:15 [FREE Full text] [doi: 10.1186/s40900-018-0098-y] [Medline: 29785283]
70. Mathie E, Smeeton N, Munday D, Rhodes G, Wythe H, Jones J. The role of patient and public involvement leads in facilitating feedback: "invisible work". Res Involv Engagem 2020;6:40 [FREE Full text] [doi: 10.1186/s40900-020-00209-2] [Medline: 32676199]
71. McCarron TL, Noseworthy T, Moffat K, Wilkinson G, Zelinsky S, White D, et al. A co-designed framework to support and sustain patient and family engagement in health-care decision making. Health Expect 2020 Aug;23(4):825-836 [FREE Full text] [doi: 10.1111/hex.13054] [Medline: 32337836]
72. Murta FR, Waxman J, Skilton A, Wickwar S, Bonstein K, Cable R, et al. The first UK national blepharospasm patient and public involvement day; identifying priorities. Orbit 2020 Aug;39(4):233-240. [doi: 10.1080/01676830.2019.1657469] [Medline: 31466502]
73. Ní Shé E, Cassidy J, Davies C, De Brún A, Donnelly S, Dorris E, et al. Minding the gap: identifying values to enable public and patient involvement at the pre-commencement stage of research projects. Res Involv Engagem 2020;6:46 [FREE Full text] [doi: 10.1186/s40900-020-00220-7] [Medline: 32765898]
74. Nissen ER, Bregnballe V, Mehlsen MY, Muldbjerg A, O'Connor M, Lomborg KE. Patient involvement in the development of a psychosocial cancer rehabilitation intervention: evaluation of a shared working group with patients and researchers. Res Involv Engagem 2018;4:24 [FREE Full text] [doi: 10.1186/s40900-018-0106-2] [Medline: 30123530]
75. O'Donnell D, Ní Shé E, McCarthy M, Thornton S, Doran T, Smith F, et al. Enabling public, patient and practitioner involvement in co-designing frailty pathways in the acute care setting. BMC Health Serv Res 2019 Nov 05;19(1):797 [FREE Full text] [doi: 10.1186/s12913-019-4626-8] [Medline: 31690304]
76. Porcheret M, Grime J, Main C, Dziedzic K. Developing a model osteoarthritis consultation: a Delphi consensus exercise. BMC Musculoskelet Disord 2013 Jan 16;14:25 [FREE Full text] [doi: 10.1186/1471-2474-14-25] [Medline: 23320630]
77. Rankin G, Summers R, Cowan K, Barker K, Button K, Carroll SP, James Lind Alliance (JLA) Physiotherapy Priority Setting Partnership (PSP) Steering Group. Identifying priorities for physiotherapy research in the UK: the James Lind Alliance physiotherapy priority setting partnership. Physiotherapy 2020 Jun;107:161-168. [doi: 10.1016/j.physio.2019.07.006] [Medline: 32026816]
78. Read S, Aries AM, Ashby SM, Bambrick V, Blackburn SJ, Clifford H, et al. Facilitating personal development for public involvement in health-care education and research: a co-produced pilot study in one UK higher education institute. Health Expect 2020 Oct;23(5):1191-1201 [FREE Full text] [doi: 10.1111/hex.13097] [Medline: 32707602]
79. Robinson T, Skouteris H, Burns P, Melder A, Bailey C, Croft C, et al. Flipping the paradigm: a qualitative exploration of research translation centres in the United Kingdom and Australia. Health Res Policy Syst 2020 Sep 29;18(1):111 [FREE Full text] [doi: 10.1186/s12961-020-00622-9] [Medline: 32993658]
80. Walsh NM, O'Brien EC, Geraghty AA, Byrne DF, Whelan A, Reilly S, et al. Taking guidance from parents involved in a longitudinal birth cohort - the ROLO family advisory committee. Res Involv Engagem 2020;6:16 [FREE Full text] [doi: 10.1186/s40900-020-00200-x] [Medline: 32368350]
81. Wittmeier KD, Hobbs-Murison K, Holland C, Crawford E, Loewen H, Morris M, et al. Identifying information needs for hirschsprung disease through caregiver involvement via social media: a prioritization study and literature review. J Med Internet Res 2018 Dec 21;20(12):e297 [FREE Full text] [doi: 10.2196/jmir.9701] [Medline: 30578208]
82. Puppo C, Spire B, Morel S, Génin M, Béniguel L, Costagliola D, et al. How PrEP users constitute a community in the MSM population through their specific experience and management of stigmatization. The example of the French

ANRS-PREVENIR study. AIDS Care 2020 May;32(sup2):32-39. [doi: 10.1080/09540121.2020.1742863] [Medline: 32174136]
83. Young HM, Goodliffe S, Madhani M, Phelps K, Regen E, Locke A, et al. Co-producing progression criteria for feasibility studies: a partnership between patient contributors, clinicians and researchers. Int J Environ Res Public Health 2019 Oct 06;16(19) [FREE Full text] [doi: 10.3390/ijerph16193756] [Medline: 31590424]
84. Sin J, Henderson C, Woodham LA, Hernández A, Gillard S. A multicomponent eHealth intervention for family carers for people affected by psychosis: a coproduced design and build study. J Med Internet Res 2019 Aug 06;21(8):e14374 [FREE Full text] [doi: 10.2196/14374] [Medline: 31389333]
85. Bell T, Vat LE, McGavin C, Keller M, Getchell L, Rychtera A, et al. Co-building a patient-oriented research curriculum in Canada. Res Involv Engagem 2019 Feb 11;5(1):7 [FREE Full text] [doi: 10.1186/s40900-019-0141-7] [Medline: 30788148]
86. Tapp H, Derkowski D, Calvert M, Welch M, Spencer S. Patient perspectives on engagement in shared decision-making for asthma care. Fam Pract 2017 Jun 01;34(3):353-357 [FREE Full text] [doi: 10.1093/fampra/cmw122] [Medline: 28034918]
87. Locock L, Kirkpatrick S, Brading L, Sturmey G, Cornwell J, Churchill N, et al. Involving service users in the qualitative analysis of patient narratives to support healthcare quality improvement. Res Involv Engagem 2019;5:1 [FREE Full text] [doi: 10.1186/s40900-018-0133-z] [Medline: 30788147]
88. Buffel T. Social research and co-production with older people: developing age-friendly communities. J Aging Stud 2018 Mar;44:52-60 [FREE Full text] [doi: 10.1016/j.jaging.2018.01.012] [Medline: 29502790]
89. Carr EC, Patel JN, Ortiz MM, Miller JL, Teare SR, Barber CE, et al. Co-design of a patient experience survey for arthritis central intake: an example of meaningful patient engagement in healthcare design. BMC Health Serv Res 2019 Jun 04;19(1):355 [FREE Full text] [doi: 10.1186/s12913-019-4196-9] [Medline: 31164176]
90. Dahm MR, Brown A, Martin D, Williams M, Osborne B, Basseal J, et al. Interaction and innovation: practical strategies for inclusive consumer-driven research in health services. BMJ Open 2019 Dec 16;9(12):e031555 [FREE Full text] [doi: 10.1136/bmjopen-2019-031555] [Medline: 31848163]
91. Dias S, Gama A, Simões D, Mendão L. Implementation process and impacts of a participatory HIV research project with key populations. Biomed Res Int 2018;2018:5845218 [FREE Full text] [doi: 10.1155/2018/5845218] [Medline: 29955605]
92. Fonseka TM, Pong JT, Kcomt A, Kennedy SH, Parikh SV. Collaborating with individuals with lived experience to adapt CANMAT clinical depression guidelines into a patient treatment guide: the CHOICE-D co-design process. J Eval Clin Pract 2020 Aug;26(4):1259-1269. [doi: 10.1111/jep.13308] [Medline: 31729117]
93. Frith L, Hepworth L, Lowers V, Joseph F, Davies E, Gabbay M. Role of public involvement in the Royal College of Physicians' Future Hospitals healthcare improvement programme: an evaluation. BMJ Open 2019 Sep 12;9(9):e027680. [doi: 10.1136/bmjopen-2018-027680] [Medline: 31515414]
94. King C, Gillard S. Bringing together coproduction and community participatory research approaches: using first person reflective narrative to explore coproduction and community involvement in mental health research. Health Expect 2019 Aug;22(4):701-708 [FREE Full text] [doi: 10.1111/hex.12908] [Medline: 31187556]
95. Leslie M, Khayatzadeh-Mahani A, MacKean G. Recruitment of caregivers into health services research: lessons from a user-centred design study. Res Involv Engagem 2019;5:17 [FREE Full text] [doi: 10.1186/s40900-019-0150-6] [Medline: 31139432]
96. McCarron TL, Noseworthy T, Moffat K, Wilkinson G, Zelinsky S, White D, et al. Understanding the motivations of patients: a co-designed project to understand the factors behind patient engagement. Health Expect 2019 Aug;22(4):709-720 [FREE Full text] [doi: 10.1111/hex.12942] [Medline: 31379094]
97. Taylor RM, Mohain J, Gibson F, Solanki A, Whelan J, Fern LA. Novel participatory methods of involving patients in research: naming and branding a longitudinal cohort study, BRIGHTLIGHT. BMC Med Res Methodol 2015 Mar 14;15:20 [FREE Full text] [doi: 10.1186/s12874-015-0014-1] [Medline: 25888163]
98. Woods L, Roehrer E, Duff J, Walker K, Cummings E. Co-design of a mobile health app for heart failure: perspectives from the team. Stud Health Technol Inform 2019 Aug 08;266:183-188. [doi: 10.3233/SHTI190792] [Medline: 31397321]
99. Tsianakas V, Maben J, Wiseman T, Robert G, Richardson A, Madden P, et al. Using patients' experiences to identify priorities for quality improvement in breast cancer care: patient narratives, surveys or both? BMC Health Serv Res 2012 Aug 22;12(1):271 [FREE Full text] [doi: 10.1186/1472-6963-12-271] [Medline: 22913525]
100. Cheraghi-Sohi S, Hole AR, Mead N, McDonald R, Whalley D, Bower P, et al. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann Fam Med 2008;6(2):107-115 [FREE Full text] [doi: 10.1370/afm.816] [Medline: 18332402]
101. Kalot MA, Al-Khatib M, Connell NT, Flood V, Brignardello-Petersen R, James P, VWD working group. An international survey to inform priorities for new guidelines on von Willebrand disease. Haemophilia 2020 Jan 26;26(1):106-116 [FREE Full text] [doi: 10.1111/hae.13881] [Medline: 31769905]
102. Brooks H, Irmansyah I, Susanti H, Utomo B, Prawira B, Iskandar L, et al. Evaluating the acceptability of a co-produced and co-delivered mental health public engagement festival: Mental Health Matters, Jakarta, Indonesia. Res Involv Engagem 2019;5:25 [FREE Full text] [doi: 10.1186/s40900-019-0161-3] [Medline: 31516732]

103. Jilka S, Murray C, Wieczorek A, Griffiths H, Wykes T, McShane R. Exploring patients' and carers' views about the clinical use of ketamine to inform policy and practical decisions: mixed-methods study. BJPsych Open 2019 Jul 30;5(5):e62 [FREE Full text] [doi: 10.1192/bjo.2019.52] [Medline: 31530293]
104. Troya MI, Dikomitis L, Babatunde OO, Bartlam B, Chew-Graham CA. Understanding self-harm in older adults: a qualitative study. EClinicalMedicine 2019 Jul;12:52-61 [FREE Full text] [doi: 10.1016/j.eclinm.2019.06.002] [Medline: 31388663]
105. Bayliss K, Starling B, Raza K, Johansson EC, Zabalan C, Moore S, et al. Patient involvement in a qualitative meta-synthesis: lessons learnt. Res Involv Engagem 2016 May 12;2(1):18 [FREE Full text] [doi: 10.1186/s40900-016-0032-0] [Medline: 29062519]
106. Best P, Badham J, Corepal R, O'Neill RF, Tully MA, Kee F, et al. Network methods to support user involvement in qualitative data analyses: an introduction to Participatory Theme Elicitation. Trials 2017 Nov 23;18(1):559 [FREE Full text] [doi: 10.1186/s13063-017-2289-5] [Medline: 29169378]
107. Garfield S, Jheeta S, Husson F, Jacklin A, Bischler A, Norton C, et al. Lay involvement in the analysis of qualitative data in health services research: a descriptive study. Res Involv Engagem 2016;2:29 [FREE Full text] [doi: 10.1186/s40900-016-0041-z] [Medline: 29507764]
108. Jennings H, Slade M, Bates P, Munday E, Toney R. Best practice framework for Patient and Public Involvement (PPI) in collaborative data analysis of qualitative mental health research: methodology development and refinement. BMC Psychiatry 2018 Jun 28;18(1):213 [FREE Full text] [doi: 10.1186/s12888-018-1794-8] [Medline: 29954373]
109. Stevenson M, Taylor BJ. Involving individuals with dementia as co-researchers in analysis of findings from a qualitative study. Dementia (London) 2019 Feb;18(2):701-712. [doi: 10.1177/1471301217690904] [Medline: 28133983]
110. Thomas F, Hansford L, Wyatt K, Byng R, Coombes K, Finch J, et al. An engaged approach to exploring issues around poverty and mental health: a reflective evaluation of the research process from researchers and community partners involved in the DeStress study. Health Expect 2020 May 24. [doi: 10.1111/hex.13065] [Medline: 32449304]
111. Lee C, Tanna N, Blair M, Yusuf Y, Khalief H, Lakhanpaul M. Getting underneath the skin: a community engagement event for optimal vitamin D status in an 'easily overlooked' group. Health Expect 2019 Dec;22(6):1322-1330 [FREE Full text] [doi: 10.1111/hex.12978] [Medline: 31605450]
112. Kuluski K, Ho JW, Cadel L, Shearkhani S, Levy C, Marcinow M, et al. An alternate level of care plan: co-designing components of an intervention with patients, caregivers and providers to address delayed hospital discharge challenges. Health Expect 2020 Oct;23(5):1155-1165 [FREE Full text] [doi: 10.1111/hex.13094] [Medline: 32602628]
113. Maccarthy J, Guerin S, Wilson AG, Dorris ER. Facilitating public and patient involvement in basic and preclinical health research. PLoS One 2019;14(5):e0216600 [FREE Full text] [doi: 10.1371/journal.pone.0216600] [Medline: 31086377]
114. Garfield S, Jheeta S, Jacklin A, Bischler A, Norton C, Franklin BD. Patient and public involvement in data collection for health services research: a descriptive study. Res Involv Engagem 2015;1:8 [FREE Full text] [doi: 10.1186/s40900-015-0006-7] [Medline: 29062497]
115. Hoddinott P, Pollock A, O'Cathain A, Boyer I, Taylor J, MacDonald C, et al. How to incorporate patient and public perspectives into the design and conduct of research. F1000Res 2018;7:752 [FREE Full text] [doi: 10.12688/f1000research.15162.1] [Medline: 30364075]
116. Morris M, Alencar Y, Rachet B, Stephens R, Coleman MP. Fleshing out the data: when epidemiological researchers engage with patients and carers. Learning lessons from a patient involvement activity. BMJ Open 2020 Sep 30;10(9):e036311 [FREE Full text] [doi: 10.1136/bmjopen-2019-036311] [Medline: 32998916]
117. Chouvarda I, Maramis C, Livitckaia K, Trajkovik V, Burmaoglu S, Belani H, ENJECT Working Group 1 Network. Connected health services: framework for an impact assessment. J Med Internet Res 2019 Sep 03;21(9):e14005 [FREE Full text] [doi: 10.2196/14005] [Medline: 31482857]
118. Chapman A, Hadfield M, Chapman C. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. J R Coll Physicians Edinb 2015;45(3):201-205. [doi: 10.4997/jrcpe.2015.305]
119. Ratzki-Leewing A, Parvaresh Rizi E, Harris SB. Family members: the forgotten players in the Diabetes Care Team (The TALK-HYPO Study). Diabetes Ther 2019 Dec;10(6):2305-2311 [FREE Full text] [doi: 10.1007/s13300-019-00687-y] [Medline: 31502110]
120. Carlin CS, Christianson JB, Keenan P, Finch M. Chronic illness and patient satisfaction. Health Serv Res 2012 Dec;47(6):2250-2272 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01412.x] [Medline: 22515159]
121. Tran V, Messou E, Djima MM, Ravaud P, Ekouevi DK. Patients' perspectives on how to decrease the burden of treatment: a qualitative study of HIV care in sub-Saharan Africa. BMJ Qual Saf 2019 Apr;28(4):266-275 [FREE Full text] [doi: 10.1136/bmjqs-2017-007564] [Medline: 29706594]
122. Morgan DL. Focus Groups as Qualitative Research. Thousand Oaks, CA: SAGE Publications Ltd; 1997.
123. Longhurst R. Semi-structured interviews and focus groups. Key Methods in Geography 2003;3(2):143-156 [FREE Full text]
124. Wiig S, Storm M, Aase K, Gjestsen MT, Solheim M, Harthug S, QUASER team. Investigating the use of patient involvement and patient experience in quality improvement in Norway: rhetoric or reality? BMC Health Serv Res 2013 Jun 06;13(1):206 [FREE Full text] [doi: 10.1186/1472-6963-13-206] [Medline: 23742265]

125. Newcomer K, Hatry H, Wholey J. Conducting semi-structured interviews. In: Handbook of Practical Program Evaluation. Hoboken, New Jersey, United States: John Wiley & Sons; 2015.
126. Cowan K. The James Lind Alliance Guidebook. National Institute for Health Research. 2018. URL: https://www.jla.nihr.ac.uk/jla-guidebook/ [accessed 2021-12-07]
127. Patient and Public Involvement (PPI) in a digital age : an overview of AHSN activity and learning. Academic Health Science Networks. 2018. URL: https://www.ahsnnetwork.com/app/uploads/2018/12/Patient-and-public-involvement-PPI-in-a-digital-world-May-2018.pdf [accessed 2021-12-07]
128. Créquit P, Mansouri G, Benchoufi M, Vivot A, Ravaud P. Mapping of crowdsourcing in health: systematic review. J Med Internet Res 2018 May 15;20(5):e187 [FREE Full text] [doi: 10.2196/jmir.9330] [Medline: 29764795]
129. Truitt AR, Monsell SE, Avins AL, Nerenz DR, Lawrence SO, Bauer Z, et al. Prioritizing research topics: a comparison of crowdsourcing and patient registry. Qual Life Res 2018 Jan 5;27(1):41-50. [doi: 10.1007/s11136-017-1566-9] [Medline: 28382522]
130. Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, et al. PREDOSE: a semantic web platform for drug abuse epidemiology using social media. J Biomed Inform 2013 Dec;46(6):985-997 [FREE Full text] [doi: 10.1016/j.jbi.2013.07.007] [Medline: 23892295]
131. Ampofo L, Collister S, O'Loughlin B, Chadwick A. Text mining and social media: when quantitative meets qualitative and software meets people. In: Halfpenny P, Procter R, editors. Innovations in Digital Research Methods. Thousand Oaks, CA: SAGE Publications Ltd; 2015:161-192.
132. Design together, live better. The AHSN Network. URL: https://www.weahsn.net/design-together-live-better/ [accessed 2020-11-20]
133. Melton H, Meader N, Dale H, Wright K, Jones-Diette J, Temple M, et al. Interventions for adults with a history of complex traumatic events: the INCiTE mixed-methods systematic review. Health Technol Assess 2020 Sep;24(43):1-312 [FREE Full text] [doi: 10.3310/hta24430] [Medline: 32924926]
134. Brooks H, Irmansyah I, Lovell K, Savitri I, Utomo B, Prawira B, et al. Improving mental health literacy among young people aged 11-15 years in Java, Indonesia: co-development and feasibility testing of a culturally-appropriate, user-centred resource (IMPeTUs) - a study protocol. BMC Health Serv Res 2019 Jul 12;19(1):484 [FREE Full text] [doi: 10.1186/s12913-019-4328-2] [Medline: 31299968]
135. Lavallee DC, Lee JR, Semple JL, Lober WB, Evans HL. Engaging patients in co-design of mobile health tools for surgical site infection surveillance: implications for research and implementation. Surg Infect (Larchmt) 2019 Oct;20(7):535-540 [FREE Full text] [doi: 10.1089/sur.2019.148] [Medline: 31429644]
136. Frisch N, Atherton P, Doyle-Waters MM, MacLeod ML, Mallidou A, Sheane V, et al. Patient-oriented research competencies in health (PORCH) for researchers, patients, healthcare providers, and decision-makers: results of a scoping review. Res Involv Engagem 2020;6:4 [FREE Full text] [doi: 10.1186/s40900-020-0180-0] [Medline: 32055415]
137. Price A, Vasanthan L, Clarke M, Liew SM, Brice A, Burls A. SMOOTH: Self-Management of Open Online Trials in Health analysis found improvements were needed for reporting methods of internet-based trials. J Clin Epidemiol 2019 Jan;105:27-39. [doi: 10.1016/j.jclinepi.2018.08.017] [Medline: 30171901]
138. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014 Feb 26;14(1):89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
139. Ludwig C, Graham ID, Gifford W, Lavoie J, Stacey D. Partnering with frail or seriously ill patients in research: a systematic review. Res Involv Engagem 2020;6:52 [FREE Full text] [doi: 10.1186/s40900-020-00225-2] [Medline: 32944284]
140. Lucas PJ, Allnock D, Jessiman T. How are European birth-cohort studies engaging and consulting with young cohort members? BMC Med Res Methodol 2013 Apr 11;13:56 [FREE Full text] [doi: 10.1186/1471-2288-13-56] [Medline: 23578172]
141. Mathie E, Wythe H, Munday D, Millac P, Rhodes G, Roberts N, et al. Reciprocal relationships and the importance of feedback in patient and public involvement: a mixed methods study. Health Expect 2018 Oct;21(5):899-908 [FREE Full text] [doi: 10.1111/hex.12684] [Medline: 29654644]
142. Chambers E, Gardiner C, Thompson J, Seymour J. Patient and carer involvement in palliative care research: an integrative qualitative evidence synthesis review. Palliat Med 2019 Sep;33(8):969-984 [FREE Full text] [doi: 10.1177/0269216319858247] [Medline: 31250702]
143. 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]
144. 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;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
145. Liabo K, Boddy K, Bortoli S, Irvine J, Boult H, Fredlund M, et al. Public involvement in health research: what does 'good' look like in practice? Res Involv Engagem 2020;6:11 [FREE Full text] [doi: 10.1186/s40900-020-0183-x] [Medline: 32266085]

146. Nunn JS, Tiller J, Fransquet P, Lacaze P. Public involvement in global genomics research: a scoping review. Front Public Health 2019;7:79 [FREE Full text] [doi: 10.3389/fpubh.2019.00079] [Medline: 31024880]
147. Miah J, Dawes P, Edwards S, Leroi I, Starling B, Parsons S. Patient and public involvement in dementia research in the European Union: a scoping review. BMC Geriatr 2019 Aug 14;19(1):220 [FREE Full text] [doi: 10.1186/s12877-019-1217-9] [Medline: 31412788]
148. Hutchison K, Rogers W, Entwistle VA. Addressing deficits and injustices: the potential epistemic contributions of patients to research. Health Care Anal 2017 Dec 8;25(4):386-403. [doi: 10.1007/s10728-016-0323-5] [Medline: 27277736]
149. Buck D, Gamble C, Dudley L, Preston J, Hanley B, Williamson PR, EPIC Patient Advisory Group. From plans to actions in patient and public involvement: qualitative study of documented plans and the accounts of researchers and patients sampled from a cohort of clinical trials. BMJ Open 2014 Dec 04;4(12):e006400 [FREE Full text] [doi: 10.1136/bmjopen-2014-006400] [Medline: 25475243]
150. Naidoo N, Nguyen VT, Ravaud P, Young B, Amiel P, Schanté D, et al. The research burden of randomized controlled trial participation: a systematic thematic synthesis of qualitative evidence. BMC Med 2020 Jan 20;18(1):6 [FREE Full text] [doi: 10.1186/s12916-019-1476-5] [Medline: 31955710]
151. Bereczky T. Patient advocacy -with a feeling patient citizenship -the description of an affective model of patient advocacy. Doctoral Dissertation, ELTE University, Hungary. 2019. URL: https://www.academia.edu/42773922/Patient_Advocacy_With_a_Feeling_Patient_citizenship_The_description_of_an_affective_model_of_patient_advocacy [accessed 2021-12-07]
152. Sacristan JA, Aguaron A, Avendaño C, Garrido P, Carrion J, Gutierrez A, et al. Patient involvement in clinical research: why, when, and how. Patient Prefer Adher 2016 Apr:631. [doi: 10.2147/ppa.s104259]
153. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
154. Algeo N, Hunter D, Cahill A, Dickson C, Adams J. Usability of a digital self-management website for people with osteoarthritis: a UK patient and public involvement study. Int J Ther Rehabil 2017 Feb 02;24(2):78-82. [doi: 10.12968/ijtr.2017.24.2.78]

Abbreviations

GRIPP2: revised version of Guidance for Reporting Involvement of Patients and the Public
PPI: patient and public involvement

Edited by R Kukafka; submitted 13.11.20; peer-reviewed by Z Aghaei, A Khaleghi; comments to author 17.12.20; revised version received 16.01.21; accepted 08.10.21; published 23.12.21.

Please cite as:
Aguayo GA, Goetzinger C, Scibilia R, Fischer A, Seuring T, Tran VT, Ravaud P, Bereczky T, Huiart L, Fagherazzi G
Methods to Generate Innovative Research Ideas and Improve Patient and Public Involvement in Modern Epidemiological Research: Review, Patient Viewpoint, and Guidelines for Implementation of a Digital Cohort Study
J Med Internet Res 2021;23(12):e25743
URL: https://www.jmir.org/2021/12/e25743
doi:10.2196/25743
PMID:34941554

Review

Genomic Health Literacy Interventions in Pediatrics: Scoping Review

Aarushi Gupta[1,2], BSc; Joseph A Cafazzo[1,2,3], BASc, MHSc, PEng, PhD; Maarten J IJzerman[4,5], PhD; Joost F Swart[6,7], MD, PhD; Sebastiaan Vastert[6,7], MD, PhD; Nico M Wulffraat[6,7], MD, PhD; Susanne Benseler[8,9,10], MD, PhD; Deborah Marshall[9,11,12], PhD; Rae Yeung[13,14], MD, PhD; Marinka Twilt[8,9,10], MD, MSCE, PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre of Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[3]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[4]Department of Health and Technology and Services Research, Faculty of Behavioural, Management and Social Sciences, Technical Medical Centre, University of Twente, Twente, Netherlands

[5]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[6]Division of Pediatrics, Department of Pediatric Rheumatology and Immunology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, Netherlands

[7]Faculty of Medicine, Utrecht University, Utrecht, Netherlands

[8]Division of Rheumatology, Department of Pediatrics, Alberta Children's Hospital, Calgary, AB, Canada

[9]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada

[10]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[11]Department of Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[12]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[13]Division of Rheumatology, Department of Pediatrics, The Hospital for Sick Children, Toronto, AB, Canada

[14]Immunology and Institute of Medical Science, University of Toronto, Toronto, AB, Canada

Corresponding Author:
Marinka Twilt, MD, MSCE, PhD
Division of Rheumatology, Department of Pediatrics
Alberta Children's Hospital
28 Oki Drive NW
Calgary, AB, T3B 6A8
Canada
Phone: 1 403 955 7771
Fax: 1 403 955 7649
Email: marinka.twilt@ahs.ca

Abstract

Background: The emergence of genetic and genomic sequencing approaches for pediatric patients has raised questions about the genomic health literacy levels, attitudes toward receiving genomic information, and use of this information to inform treatment decisions by pediatric patients and their parents. However, the methods to educate pediatric patients and their parents about genomic concepts through digital health interventions have not been well-established.

Objective: The primary objective of this scoping review is to investigate the current levels of genomic health literacy and the attitudes toward receiving genomic information among pediatric patients and their parents. The secondary aim is to investigate patient education interventions that aim to measure and increase genomic health literacy among pediatric patients and their parents. The findings from this review will be used to inform future digital health interventions for patient education.

Methods: A scoping review using PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines and protocols was completed using the following databases: MEDLINE, Embase, CINAHL, and Scopus. Our search strategy included genomic information inclusive of all genetic and genomic terms, pediatrics, and patient education. Inclusion criteria included the following: the study included genetic, genomic, or a combination of genetic and genomic information; the study population was pediatric (children and adolescents <18 years) and parents of patients with pediatric illnesses or only parents of patients with pediatric illnesses; the study included an assessment of the knowledge, attitudes,

and intervention regarding genomic information; the study was conducted in the last 12 years between 2008 and 2020; and the study was in the English language. Descriptive data regarding study design, methodology, disease population, and key findings were extracted. All the findings were collated, categorized, and reported thematically.

Results: Of the 4618 studies, 14 studies (n=6, 43% qualitative, n=6, 43% mixed methods, and n=2, 14% quantitative) were included. Key findings were based on the following 6 themes: knowledge of genomic concepts, use of the internet and social media for genomic information, use of genomic information for decision-making, hopes and attitudes toward receiving genomic information, experiences with genetic counseling, and interventions to improve genomic knowledge.

Conclusions: This review identified that older age is related to the capacity of understanding genomic concepts, increased genomic health literacy levels, and the perceived ability to participate in decision-making related to genomic information. In addition, internet-searching plays a major role in obtaining genomic information and filling gaps in communication with health care providers. However, little is known about the capacity of pediatric patients and their parents to understand genomic information and make informed decisions based on the genomic information obtained. More research is required to inform digital health interventions and to leverage the leading best practices to educate these genomic concepts.

(J Med Internet Res 2021;23(12):e26684) doi:10.2196/26684

KEYWORDS

pediatrics; patient education; genetics; genomics; mHealth; digital health; internet; genetic knowledge; genomic health literacy; children; adolescents

Introduction

Background

Recent scientific breakthroughs and technological advancements in personalized and precision medicine are changing the way we diagnose and treat diseases, leading to more precise, predictable, and powerful health care that is customized for the individual patient [1]. However, individualized diagnostic and treatment pathway development is expensive and introduces new aspects of patient engagement into the more traditional medical practice. Personalized and precision medicine and genome sequencing have gone hand in hand and become more widely available and incorporated into clinical pediatric and adolescent care, either in the context of routine patient care or research [2,3]. However, genetic counselors have indicated that they lack the relevant knowledge, confidence, and practical techniques to educate adolescents about genomic concepts [4]; some health professionals have also expressed uncertainty about the cognitive abilities of adolescents to understand genomic concepts [5].

Genomic health literacy is defined as the basic knowledge of genetic and genomic concepts and the capacity to obtain, process, understand, and use genomic information for health-related decision-making [6]. Studies have shown that children and adolescents have the desire to learn more about the genetic factors related to their illness and to be more involved in the decision-making process of their treatment [7-11]. Moreover, pediatric patients undergoing genomic sequencing and their parents have expressed the desire to learn about actionable genomic research results [12-14]. As such, the increasing number of clinical genetic tests, research endeavors that use exome and genome sequencing, and increasing professional opportunities in genomics (eg, bioinformatics and genetic counseling) for adolescents entering the workforce point to a need to develop educational material on genomics for young people [15]. A systematic review by McGill et al [7] found that although children and adolescents in the general community may have a basic understanding of genetic concepts such as inheritance, they generally lack a deeper knowledge of concepts related to genetics and genetic testing. Although a high level of genomic health literacy is unlikely in children and adolescents, it may be valuable for young people who are affected or at risk of genetic conditions to have a general understanding of genomic concepts [7].

Navigating through the transitional stages of childhood and adolescence with a genetic condition could lead to difficulties with autonomy, identity development, and self-esteem [16]. Moreover, results from genetic testing of a child may have implications for parents and other family members [17]. The American College of Medical Genetics and Genomics guidelines recommend that children as young as 8 years should be actively involved in the decision and interpretation of the clinical exome or genome sequencing process to the extent that they are considered cognitively capable, which includes the assent of the child whenever reasonable and respecting their preferences [18].

Research has highlighted both positive and negative implications for the psychological outcomes for those who undergo genetic testing [19], and the ethical implications for returning genetic information to children have been widely debated, especially in the context of informed consent for genetic and genomic testing [20]. This attention resonates with the presumptions that parents know what is in their children's best interest and that minors are unable to provide informed consent [20,21]. As children age, they gain decision-making capacity and an understanding of health conditions [20]. Therefore, including children and adolescents to various degrees as they age in health decisions related to genomic information is important yet challenging.

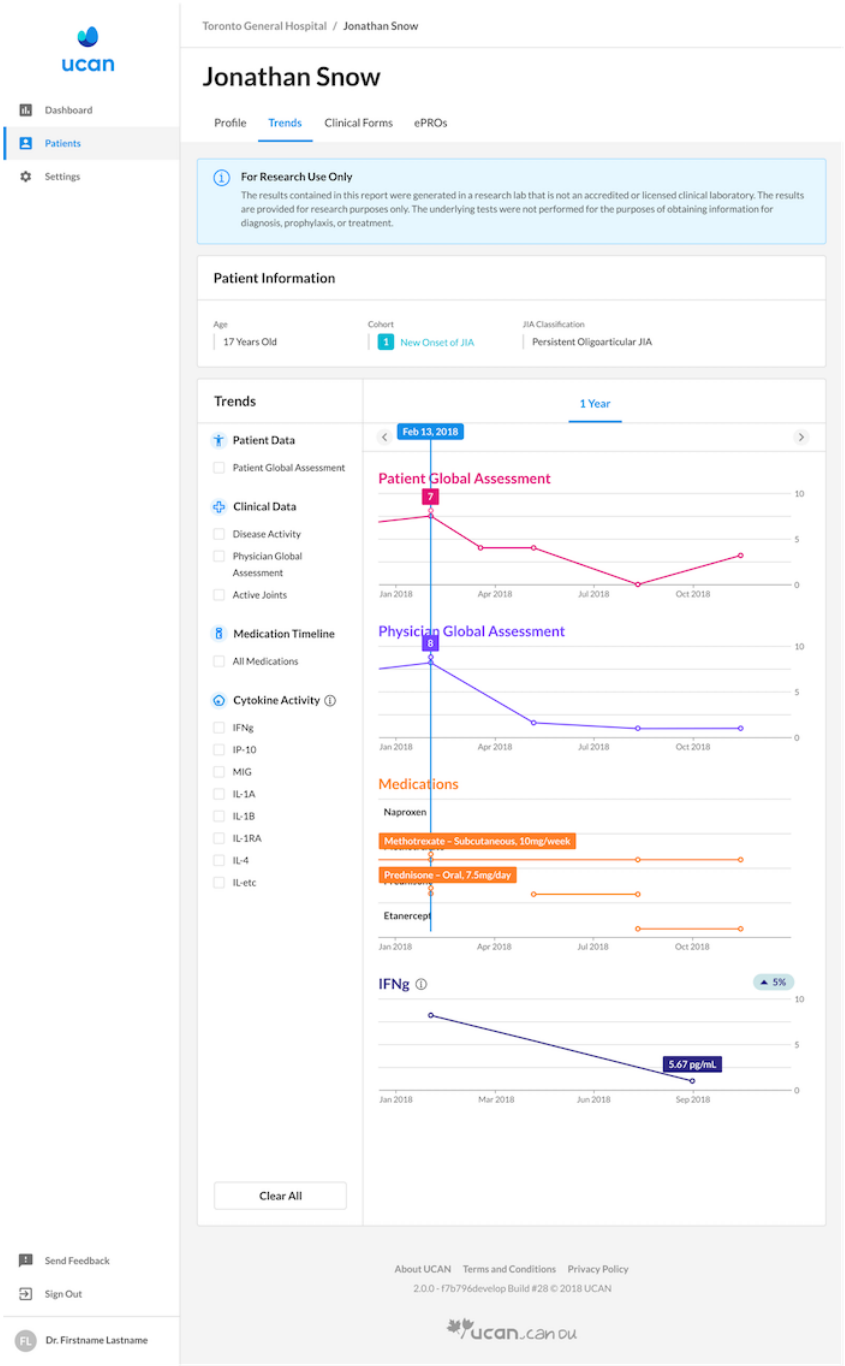
Goal of the Study

This review was conducted as formative research for the Understanding Childhood Arthritis Network (UCAN) team to inform the design of a digital health intervention: a genomics patient education feature for patients living with juvenile idiopathic arthritis (JIA). JIA is the most common childhood

chronic rheumatic disease and has a prevalence of 16-150 cases per 100,000 population [22] and can have a negative impact on the health-related aspects of quality of life [23]. Children and adolescents with JIA experience physical symptoms such as stiffness, fatigue, and sleep impairments; emotional symptoms such as stress, anxiety, and depression; and reduced social interactions [23]. As the etiology of JIA is unknown and is currently attributed to different genetic and environmental factors, a variety of pharmacological therapies are used to manage symptoms [24]. Biological disease-modifying antirheumatic therapies that target specific cytokines involved in the inflammatory cascade, such as tumor necrosis factor-α inhibitors, interleukin-1 inhibitors, and interleukin-6 inhibitors, have greatly changed the outcomes and morbidity associated with JIA but are associated with high costs [25,26] and safety concerns such as the risk of infection [25,26].

The ongoing UCAN study combines genomic discovery with patient-reported outcomes and health economic analyses to identify children at high risk of poor disease outcomes, define optimal ways to manage affected children, and develop a sustainable transdisciplinary network to improve the quality of life for all children with arthritis. One of the key features of the innovative UCAN platform is a novel genomics dashboard, which displays genomic information and trends in cytokine activity for patients. This tool acts as a visual aid for providers to discuss the severity of childhood arthritis with the patients with JIA and their parents and to identify the potential treatment targets based on the patient's genomic profile (Figure 1).

Figure 1. Understanding Childhood Arthritis Network platform genomics dashboard.



In studies such as UCAN, where genomic information is shared with patients, there is a need to educate patients and their parents by improving the levels of genomic health literacy; to foster a better understanding of the disease; for meaningful conversation; for decision-making for disease management; and to foster a better understanding of implications on treatment, outcomes, quality of life, and long-term consequences. Although numerous studies have explored genomic health literacy among adult populations [27-29] and interventions to improve patient genetic education [30], there is a general lack of research regarding how children and adolescents understand genetic illnesses [7] and genetic and genomic testing, how such information should be conveyed to them, and what factors may affect communication efficacy [8]. In addition, no studies have investigated genomic health literacy or tools to educate the patients with JIA and their parents about genomic concepts.

Objectives

We aim to perform a scoping review to identify and synthesize existing literature regarding the genomics health literacy levels and attitudes relative to receiving genomic information among pediatric patients and their parents to identify current practices and existing interventions that aim to improve genomic health literacy among pediatric patients and their parents. The term *genomics* is used as an umbrella term throughout the review as it encompasses the fields of genetic and genomic information; genomics describes the study of genes in their entirety, including their function, interaction, and environment and application of genome-based strategies [31]. Our research questions were intentionally broad to capture all the relevant literature relating to genomic health literacy and informational needs within pediatric patients:

- What are the genomic health literacy levels and attitudes toward receiving genomic information among pediatric patients and their parents?
- What interventions are known to educate pediatric patients and their parents about genomics to improve genomic health literacy levels and to facilitate a better understanding of their treatments?

Methods

Protocol and Registration

The protocol was not registered as scoping review protocols do not require registration. The scoping review methodology used was modeled based on the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) protocol [32]. This approach was used to study all the aspects of the topic, allow for a comprehensive exploration of patient knowledge and experiences, identify the existing literature relevant to the topics of interest, and identify gaps in the evidence.

Eligibility Criteria

Studies were included if they met all of the following eligibility criteria: the study included genetic, genomic, or a combination of genetic and genomic information; the study population was pediatric (children and adolescents <18 years) and parents of pediatric patients or only parents of pediatric patients; the study included an assessment of the knowledge, attitudes, and understanding of genomic information; the study was conducted in the last 12 years between January 2008 and September 2020; the study only included human participants; and the study was in the English language. The time frame of the past 12 years was selected to capture the most recent and emerging practices in the genomics field.

Information Sources

A formal electronic search and extraction was conducted between June 2020 and September 2020 on 4 electronic databases: MEDLINE, Embase, CINAHL, and Scopus. A gray literature search was also conducted on the following databases: Open Gray, CenterWatch, Cochrane Library, University of Toronto Libraries, TRIP database, ISRCTN registry, and advanced Google search.

Guidance and support from a faculty-affiliated librarian at the University of Toronto was received to formulate keywords and subject headings for the search strategy.

Search Strategy

The search strategy included 3 main concepts: pediatrics, patient education, and genomics. Textbox 1 displays the full electronic search strategy used for MEDLINE. Multimedia Appendix 1 outlines corresponding searches for all the databases used during the search process for this scoping review.

Textbox 1. MEDLINE search strategy for the scoping review on genomic knowledge and education interventions in pediatrics inform digital health interventions.

Search strategy used for MEDLINE

1. pediatric.mp. or exp Pediatrics/
2. paediatric*.tw,kf.
3. JIA.tw,kf.
4. ([pediatric or paediatric] adj5 illness).tw,kf.
5. “juvenile idiopathic arthritis”.tw,kf.
6. ([young or adolescen* or child*] adj5 illness).tw,kf.
8. autoimmune disease.mp. or exp Autoimmune Diseases/
9. Patient education.mp. or exp Patient education as Topic/
10. ([patient or young or adolescen* or child*] adj5 [educat* or learn* or knowledge or literacy or info*]).tw,kf.
11. “health knowledge”.tw,kf.
12. exp Genomics/
13. (genetic* or genomic* or genom* or biologic*).tw,kf.
14. exp Genetic Counseling/
15. (genetic adj5 [counselling or counseling]).tw,kf.
16. “genetic testing”.tw,kf.
17. “genom* sequencing”.tw,kf.
18. 1 or 2 or 3 or 4 or 5 or
19. 7 or 8 or 9
20. 10 or 11 or 12 or 13 or 14 or 15
21. 16 and 17 and 18
22. limit 19 to english
23. limit 20 to last 12 years

Selection of Sources of Evidence

Duplicates were removed electronically, and the titles and abstracts were screened by 2 reviewers (AG and MT). Related articles that were removed during the screening process were stored in a reference list for relevant studies. The full-text screening was conducted (by AG and MT), and any discrepancies and disagreements were resolved by discussion and consensus. The most common reasons for exclusion of articles were that they involved a nonpediatric population (eg, health practitioners and adult patients), did not convey any genetic or genomic information to patients, did not have human participants, were not in the English language, were conference abstracts, were not published between 2008 and 2020, and were focused on disease etiology rather than genetic or genomic information.

Data Items and Data Charting Process

To chart and extract data from the articles selected for the scoping review, an extraction criterion was developed by 2 reviewers (AG and MT) to extract information from each publication about the study country, city, urban or not urban geography, population, sample size, age of participants, pediatric disease types, duration, demographic information, design, methodology, journal of publication, and results of the study (Multimedia Appendix 2 [3,12,16,33-43]).

Synthesis of Results

A thematic analysis of the nature and content of the articles was conducted to identify the common and recurring themes, topics, ideas, and patterns and to categorize the articles [44]. Both reviewers examined the data produced from charting and data extraction and identified the key codes relating to the research questions. The codes were used to summarize and report studies according to their main findings. A total of 6 major themes were identified during the thematic analysis: *knowledge of genomic concepts*, *use of the internet and social media for genomic information*, *use of genomic information for decision-making*, *hopes and attitudes toward receiving genomic information and support*, *experiences with genetic counseling*, *and interventions to improve knowledge of genomics.*

Results

Screening Process

A total of 4583 articles were identified from 4 electronic databases listed, and 35 articles were identified from a gray literature search. After the removal of 540 sets of duplicates, the remaining 3349 articles were screened according to the

eligibility criteria. After the title and abstract screening stages, 41 articles were included in the full-text review and 27 articles were excluded owing to the reasons outlined in Multimedia Appendix 3. In total, 14 studies were included in the scoping review.

Overview of the Included Studies

Of the 14 studies, 5 (36%) studies included populations where children were clinically diagnosed with various genetic conditions [16,33-36]; 4 (29%) studies had populations that were not diagnosed with an illness [12,37-39]; 1 (7%) study included children who were suspected to have a genetic condition but not diagnosed [40]; 1 (7%) study included children who were hospitalized for various reasons [3]; and the remaining 3 (21%) studies reported pediatric populations with illnesses such as cancer [41], congenital heart defects [42], and congenital lower limb deficiencies [43]. Most (11/14, 79%) of the studies selected for the review were from the United States [3,12,33-36,38-42], 14% (2/14) from Canada [16,43] and 7% (1/14) from the United Kingdom [37]. Table 1 presents the overview of the included studies.

Table 1. Overview of the studies included in the scoping review (N=14).

Theme and study	Population	Samples, n	Study design	Aims of the research
Knowledge of genomic concepts				
Fitzgerald-Butt et al [42]	Parents of children with LVOT[a]	287	Quantitative	To examine the genetic knowledge and attitudes toward genetic testing of parents of children with heart defects affecting the LVOT
Gallo et al [34]	Parents in families in which the child has a single gene condition	142	Mixed methods	To identify unique patterns of information management and to explore the relationship between these patterns and individual and family characteristics and functioning
Lewis et al [37]	Children	539	Quantitative	To develop and validate a robust kids-KOGS[b] suitable for use in the pediatric setting and for general public education
Rew et al [39]	Parents and adolescents	33 (22 adolescents and 11 parents)	Qualitative	To determine the levels of knowledge about genetics and approaches to decision-making related to genetic testing among adolescents and parents
Use of the internet and social media for genomic information				
Barton et al [33]	Parents of children (<18 years) who underwent genetic testing	20	Qualitative	To analyze parent views about the use of the internet and social media for informational and emotional support needs at different stages of their child's genetic testing process
Roche et al [35]	Parents of children referred for genetic services	100	Qualitative	To investigate how parents of a child referred for genetic services search the internet for information before and after referral to a university pediatric genetics clinic, interpret and evaluate the information they obtained, and identify barriers that they encountered
Schaffer et al [36]	Mothers of children with genetic disorders	100	Qualitative	To investigate how mothers of children with genetic disorders use the internet to interpret, produce, and circulate genetic knowledge pertaining to their child's condition; come to value their own experiential knowledge; and help shift the boundaries of what is considered as authoritative knowledge
Use of genomics information for decision-making				
McGowan et al [12]	Parents and adolescents (aged 13-18 years)	33 (15 adolescents and 18 parents)	Qualitative	To investigate decision preferences about values and involvement in choices of genomic sequencing results and to inform and guide practices of genomic researchers working with adolescents
Myers et al [38]	Parents and adolescents (aged 13-17 years)	326 (163 dyads)	Mixed methods	To examine decisions about learning genomic research results for the adolescents and whether choices were associated with demographic factors
Hopes and attitudes toward receiving information and support				
Campbell et al [43]	Parents of children with CLD[c]	25	Mixed methods	To collect data on Canadian pediatric patients affected by CLD followed to determine emotional supports, communication information, and implementation of genetics referrals

Theme and study	Population	Samples, n	Study design	Aims of the research
Khan et al [40]	Adults and parents of children with a suspected genetic condition	270, (191 adults and 79 parents)	Mixed methods	To investigate motivation and perceived resources to predict the amount and kinds of information that adult patients and parents of pediatric patients hoped to receive from diagnostic sequencing results
Experiences with genetic counseling				
Pichini et al [16]	Adolescents	11	Qualitative	To investigate the experiences and perspectives with respect to genetic counseling interactions and to understand adolescent-specific issues to better educate and support this population of patients
Interventions to improve knowledge of genomics				
Johnson et al [41]	Parents of children enrolled in the Genomes for Kids program; patients with cancer	121	Mixed methods	To determine whether a 2-step consent using a structured communication model would improve knowledge and understanding of key genetic concepts
Newcomb et al [3]	Children (aged 5-10 years) and parents	52 (26 children and 26 parents)	Mixed methods	To determine whether an original children's book contributes to learning about the meaning of the terms *DNA* and *gene* in a sample of school-age children and whether experiencing the book with a pediatric nurse results in a better understanding of basic concepts than experiencing the book with a parent

[a]LVOT: left ventricular outflow tract.

[b]A 10-item knowledge of genome sequencing measure for young people.

[c]CLD: congenital limb deficiency.

Results of the Review

Knowledge of Genomic Concepts

Of the 14 studies, 4 (29%) studies investigated the knowledge and understanding of genetic concepts among pediatric patients or parents of pediatric patients [34,37,39,42]. These studies found varying results among participants of different age groups regarding the knowledge and understanding of genetic concepts including DNA, genome, genetic and environmental factors, the human genome project, and the sharing of genetic information.

Lewis et al [37] found that among school children between the ages of 11 and 15 years completing a 10-item kids-knowledge of genome sequencing measure for young people, the mean score was 4.24 (SD 2.49), on a scale where 0=low knowledge and 10=high knowledge. Age was also positively associated with the score in multivariate linear regression and the mean kids-knowledge of genome sequencing score was higher among girls than boys (4.44 vs 4.09, respectively; t_{535}=1.61; P<.001). The most frequent correctly answered questions by children were related to DNA, such as *Our DNA is inside our cells* and *Our DNA doesn't have an effect on how our body works* and the most frequent incorrectly answered questions were related to the genome, such as *Around 1% of our genome is the same as other people's* and *Our complete set of DNA is called our genome*.

Fitzgerald-Butt et al [42] tested genetic knowledge using a modified 18-item true or false instrument among parents of children with congenital heart defects and found that the mean genetic knowledge summary score was 73.8% correct. The most frequent correctly answered items were related to the interaction of genetic and environmental factors, such as *some diseases are caused by genes, environment, and lifestyle* (true; 97.2% correct) and *genes determine traits such as height, eye color and facial appearance* (true; 97.8% correct). The questions that the participants had the most difficulty with were related to basic genetic knowledge, such as identifying that *humans have 20 pairs of chromosomes* (false; 28% correct) and *parents pass both copies of each chromosome to their child* (false; 36.5% correct) are both false statements. Furthermore, educational attainment and household income were directly and significantly associated with genetic knowledge (P<.001).

Among parents and adolescents, Rew et al [39] found that although most participants had heard of genetic testing, the knowledge about the human genome project was generally lacking and inaccurate among younger adolescents (14-17 years), whereas older adolescents (18-21 years) demonstrated a better knowledge and accurate understanding of the human genome project. Most participants listed the internet and physicians as the sources of additional genetic information, and few participants listed books, articles, testing sites, teachers, and professional organizations as their sources.

Gallo et al [34] identified 4 unique information management patterns among parents of children (3-15 years) who have a single gene condition: *accurate understanding–open pattern* (30/86, 35%), *accurate understanding–selective pattern* (21/86, 24%), *discrepant understanding pattern* (13/86, 15%), and *confused understanding pattern* (22/86, 26%). In the accurate understanding–open and accurate understanding–selective (51/86, 59%) patterns, the parents had an accurate understanding of genetic concepts and were differentiated from one another based on their views about sharing information. The participants in the accurate understanding–open group actively sought information about conditions in addition to the information received from the health care providers (HCPs) and were open to sharing information about the child's condition. On the other hand, participants in the accurate understanding–selective group struggled with sharing information about the child's condition. In the discrepant understanding group, the parents within a family differed in the accuracy of their understanding of the genetic aspects of the condition and varied in their beliefs about seeking and sharing information. In the confused understanding group, the parents generally had an inaccurate understanding of one or more of the genetic aspects of the condition and some felt that they were unable to share information with others owing to a lack of understanding.

Use of Internet and Social Media for Genomic Information

Of the 14 studies, 3 (21%) studies investigated how parents of children with genetic disorders or children referred for genomic services searched the internet for information and emotional support about their child's condition [33,35,36]. Barton et al [33] interviewed the parents of children who underwent genetic testing for clinical care, Roche et al [35] interviewed parents of children referred for genetic services, and Schaffer et al [36] interviewed mothers of children with genetic disorders.

Barton et al [33] reported that at each stage of the genetic testing process (ie, before testing, pending results, and after results), informational and support needs of the parents were different. Before testing, many parents had little knowledge of genetic testing or conditions; some parents said that knowledge of genetic conditions and testing was restricted to Down syndrome. The internet was used to explore the possible diagnoses or explanations for their child's symptoms or challenges before testing, information about the genetic process during testing, and information about their child's new diagnosis and possible treatments after testing [33].

All 3 studies found that parents search the internet to learn about the child's condition, locate services for treatment, and find emotional support [33,35,36]. For example, Facebook groups [33], personal web pages, listservs, and chat rooms hosted by parent support groups [36] were mentioned as important resources to find support networks of families with similar experiences. Parents also found that searching using symptoms or diagnostic terms on widely available search engines such as Google or Yahoo or other websites sponsored by large health or advocacy groups (American Medical Association, Web MD, National Organization for Rare Disorders) played a key role in web-based searches [35]. Other targets for parents' searches included preparing for the visit, learning about genetic testing options, diagnostic and prognostic information, management and treatment, finding clinical trials, and reading about research advances [33,35,36].

Roche et al [35] reported that the advantages of using the internet for information included convenience, feeling that clinicians were taking the parents more seriously, privacy, and the ability to find previously unobtainable information. In addition, Schaffer et al [36] reported that internet-searching allowed parents to gain traditional forms of scientific literacy, confidence in communicating with clinicians, and a sense of authority over genetic knowledge. Barriers to using the internet for information included emotional distress, unavailability of valid diagnosis to search for, discouragement from providers, misinformation, false hope, anxiety, concerns about the child's privacy [33], keywords for searches, relevancy, and difficulty in remembering or spelling the diagnosis [35].

Use of Genomics Information for Decision-making

Of the 14 studies, 3 (21%) studies explored parents' and adolescents' preferences in decision-making in relation to genomic sequencing results [12,38,39]. These studies reported mixed feelings about receiving genomic results; some participants felt that it may be burdensome or raise privacy concerns [12], whereas others felt that it would help with future planning [12,38]. Myers et al [38] found that adolescents, in particular, expressed a desire to receive genomic information. However, adolescents were significantly less likely than parents to learn all results and that carrier status was the most frequent category that adolescents chose to learn about followed by adult-onset conditions, preventable conditions, and treatable conditions.

McGowan et al [12] found mixed perceptions among the participants regarding the adolescents' capacity to participate in decision-making regarding genetic results [12]. In general, the participants agreed that participation in decision-making about the return of genomic research results should depend on age, maturity, and personality of the adolescent [12]. Regarding collaborative decision-making, parents felt that they should have the final say in decision-making about the return of genomic results. However, many adolescents felt that their decisional preferences would differ from their parents and that a collaborative decision-making model involving a health care representative could serve as an advocate for adolescents' preferences.

Rew et al [39] found that when asked to make decisions about genetic testing, the mean age that young adolescents (aged 14-17 years) suggested was 16.6 years, whereas older adolescents (aged 18-21 years) suggested a higher mean age of 18.25 years. Almost half (45.5%) of the young adolescents also said that their parents would be their main source of information and advice related to decision-making regarding genetic testing.

Hopes and Attitudes Toward Receiving Genomic Information and Support

Of the 14 studies, 2 (14%) studies explored the hopes and attitudes toward receiving genomic information and support among parents of pediatric patients [40,43]. Khan et al [40]

investigated the types of information used by adult patients and parents of pediatric patients who have a suspected genetic condition that has not been definitively explained or diagnosed. Campbell et al [43] explored the emotional supports, communication information, and implementation of genetic referrals among parents of children with congenital heart defects. Khan et al [40] found that the most common kinds of information that parents hoped to learn from diagnostic sequencing were the cause of illness (119/269, 44.2%), directions for illness management (98/269, 36.4%), diagnosis (72/269, 26.8%), disease risk for family members (62/269, 23%), helping others (31/269, 11.5%), advancing science (16/269, 5.9%), miscellaneous knowledge (14/269, 5.2%), prevention (8/269, 3%), and family planning (5/269, 1.9%).

Campbell et al [43] found that 16.7% of parents reported they were very satisfied, 33.3% were satisfied, 25% felt neutral, and 25% felt dissatisfied with the emotional support they received from their HCPs. Moreover, 80% of parents did not recall being referred to a support group by their HCP. When asked whether their child had been given a specific diagnosis, 48% of parents could not correctly recall their child's specific diagnosis and 72% of parental classifications did not correspond to specific clinical classifications. In total, 56% of parents also reported that they sought additional information resources after talking to their HCP.

Experiences With Genetic Counseling

Of the 14 studies, 1 (7%) study investigated the experiences of genetic counseling for adolescents with a genetic condition [16]. The 3 main themes that emerged during interviews were understanding the genetic counselor's role, increasing perceived personal control, and adolescent-specific factors influencing adaptation to one's condition. The participants generally felt that they had a better understanding of what genetic counseling entailed and the distinct differences of a genetic counselor's role from other HCPs after the session. This was owing to the discussion of biological pathways, inheritance and recurrence risk for future children, and preparation for the future during the session. In addition, all the participants felt that learning about genetics, inheritance, and origins of the condition was an important outcome of the genetic counseling session. In addition, genetic counseling helped the participants to contextualize the condition as part of their identity, receive anticipatory guidance about the future, and feel a higher sense of ownership and control over their health.

The main adolescent-specific factors that were reported to influence adaptation to one's genetic condition were isolation, social connectedness, independence and privacy, and the timing of genetic counseling [16]. Adolescents noted feeling isolated as a result of their condition and being treated differently to others. In addition, the adolescents' perspectives about their condition were influenced by social connectedness with family, peer groups, and the larger community of other individuals with the same condition. As such, the participants expressed a desire to fit in and be perceived as *normal* by their friends. Although adolescents sought social connectedness, they expressed a need for independence and privacy, particularly within the family construct. Finally, the participants stated that they felt a greater sense of stability during the middle school to high school period than during the elementary school period and felt that it was a more appropriate timing for genetic counseling.

When asked about the tools and strategies for genetic counseling practice, the participants suggested using video clips and animations on a computer or tablet to describe the inheritance or biological processes, the normalization of their condition, and the ability to choose whether parents are present for all or a part of the session. In addition, the adolescents suggested that the genetic counselor could assist them with identifying reliable resources to gather information about their condition, support groups, and a postsession letter highlighting the key points relevant to them.

Interventions to Improve the Knowledge of Genomics

Of the 14 studies, 2 (14%) studies investigated the impact of interventions in improving the knowledge and understanding of genetic concepts [3,41]. The results of these 2 studies are presented in Table 2. Both studies used the Genetic Knowledge Index (GKI) developed to assess lay knowledge of genetic concepts among a general population not known to be at risk for genetic disease and not exposed to genetic counseling or research involving genomics, to test the understanding of genetic concepts among participants [45].

Table 2. Overview of studies that investigated the impact of interventions on improving the knowledge and understanding of genetic concepts.

Study	Aims of the research	Pretest results	Posttest results
Newcomb et al [3]	Whether an original children's book called "What DNA Does," designed as a visual aid to assist in the assent process for children enrolling in genetic testing research, could increase the child's and parent's understanding about "DNA" and "genes" and whether children reading the book with a pediatric nurse would result in a better understanding of genetic concepts than reading the book with a parent.	• Median GKI[a] score was 4 (0=all incorrect and 5=all correct), with 54% of respondents making only 1 incorrect response; most difficult item: "Racial differences in academic ability are caused by genetics." • Both parents' and child's understanding of the terms was minimal before reading the book; no participants mentioned learning about genetics or DNA in school. • A total of 65% (17/26) of parent respondents said they did not know what DNA was or stated vague or inaccurate definitions. • None of the child respondents was able to explain DNA in simple terms, although some were able to repeat phrases they had heard. • The primary recurring theme in the conversations about DNA before reading the book was "blood"; both children and parents expressed the idea that DNA is somehow closely related to or is a part of blood and that blood has something to do with human identity.	• GKI was not completed after the test. • After reading the book, most children had more articulate and accurate understandings of "DNA," but no better understanding of its function; 2 children were more confused after reading than before. • Children who read the book with a nurse had a better understanding of DNA's function than those who read it with a parent. • Increased accuracy of describing the meanings of DNA and gene was demonstrated by all the participants in the nurse–child–parent reading group and in two-third of the children in the parent–child reading group.
Johnson et al [41]	Whether a 2-step consent using a structured communication model would improve the knowledge and understanding of key genetic concepts among parents of children with cancer. The model involved a single study nurse who approached and obtained consents from the families with a standardized script, an informational cover sheet, and baseline pretest responses to educate parents on genetic concepts during the study introductory visit. At the subsequent informed consent visit, the nurse used a checklist and an informed consent document to review and reinforce concepts.	• More than 85% of the parents identified correct answers to 4 of 11 genetic concepts; most knew that "genes are made of DNA," "genetic risk is the chance of having an inherited (passed down) disease or disorder," "healthy parents can have a child with an inherited disease," and "genomic testing of your child's tumor and healthy tissue may teach you things about (multiple choice responses).". • Baseline understanding of differences between somatic and germline mutations was poor; 31% of parents answered correctly, "nontumor (germline) mutations are in every cell of your body," and 18% answered correctly, "tumor (somatic) mutations are only found in cancer cells."	• After completion of the 2-visit intervention, correct responses increased significantly for 9 of 11 genetic concepts and overall genetic knowledge; the median percentage of total correct answers improved from 77.8% to 88.9%. • The rate of understanding that somatic mutations are only found in cancer cells increased from 18% to 59% and understanding that germline mutations are found in every cell of the body went from 31% to 64%. • No association was detected between the change in the overall percentage of correct answers and parental numeracy, literacy, or sociodemographic factors.

[a]GKI: Genetic Knowledge Index.

Discussion

Principal Findings

This review reveals important information regarding the current genomic health literacy levels among pediatric populations and the attitudes they hold toward receiving genomic information and decision-making. The findings from this review are valuable in informing the design of digital health platforms, such as the UCAN genomics education platform, that aims to educate young patients with JIA about genomic concepts. We found that age is associated with increased genomic health literacy levels and increased perceived capacity to participate in decision-making regarding genomic information. It was also found that internet-searching is valued by young people and parents for information on diagnoses and symptom management and to fill critical gaps in communication with their HCPs. In addition, we found that adolescents found key differences in receiving genomic information from genetic counselors, compared with HCPs, which helped them contextualize their condition as a part of their identity.

Patient education regarding genomics is an emerging area of research and nearly half of the selected studies describe the goal of assessing genetic or genomic knowledge among the participants. The studies in this review found that genomic

health literacy is generally low among patients with genetic conditions and their parents. Several studies in our review found that age plays a significant role in young people's understanding of genetics [16,37,39] and their decision-making capabilities [12,38]. These findings might reflect the general teaching practices in the United States, Canada, and the United Kingdom, where genetic concepts, such as genetics, inheritance, and DNA, are formally introduced to students during the high school period (15-16 years) [46-48]. A study by Dougherty et al [46] assessing grade 12 students' understanding of essential genetic concepts across the United States, using core concepts developed by the American Society of Human Genetics as normative benchmarks, found that the states' understanding of genetic concepts was generally poor, with more than 85% of the states receiving overall scores of *inadequate*. A total of 14% (2/14) of studies in our review evaluated the change in understanding and knowledge of genetic concepts using the GKI among the participants [3,41]. These studies found that the oral presentation of information combined with a visual aid such as a brochure [41] or a children's book [3] improved understanding among participants. Both studies also indicated that the presence of a nurse to explain genetic information was beneficial in guiding knowledge acquisition among both children and parents. Although both interventions reported an improvement in genetic literacy after the intervention compared with before the intervention, the understanding of some genetic concepts such as the difference between somatic and germline mutations [41] and DNA functionality [3] remained suboptimal after the intervention. Owing to the complex nature of genetic concepts and the general difficulty that children face in understanding the functionality of genes, genetics [3,37], and whole-genome sequencing [39], it may be valuable to limit patient education to high-level information about genomics, genetics, and how they relate to the patient's disease.

Numerous studies highlighted the importance of internet-searching and seeking emotional support on the web among parents of children with genetic illnesses [33,35,36,39,45]. In general, searching the internet was reported as a key step in knowledge acquisition, improving understanding, and finding treatment options and to fill the gap of social support by finding networks of families with similar experiences. These findings are consistent with research on patients with JIA, who have listed the internet as a source for general information on JIA and emotional support [49,50]. Studies also reported that having a diagnosis played a key role in internet-searching; other research in children undergoing exome sequencing has found that families place significant value on receiving a timely diagnosis, information, and knowledge for rare illnesses [14,51]. The internet has drastically increased parents' access to the previously privileged health information, potentially changing their expectations and affecting their relationships with HCPs [36,52-54].

eHealth users list the internet as a source of health information but may not always feel comfortable sharing this information with their child's physician; a study by Tuffrey and Finlay [55] found that 84% of parents who used the internet before a pediatric visit evaluated the information they obtained as useful but only a small proportion of these parents discuss what they found on the internet with their child's physician. Similarly, several studies in this review reported that parents generally felt dissatisfied with the information they received from their health care practitioners [43], received discouragement from providers to search the internet for information, [33] or felt uncomfortable sharing the information they found on the internet with their physicians [35]. In addition, parents expressed the need for receiving more information from their HCPs and sought additional resources after talking to their HCPs. These findings are similar to the research with patients with JIA; a study by van Dijkhuizen et al [56] found that patients with JIA were most dissatisfied with the low rates of referrals and the information about immunizations, research, and existence of transition of care clinics. It is evident that eHealth resources play a large role in knowledge acquisition for parents of pediatric patients. Future research should aim to find strategies to improve knowledge-sharing among HCPs, patients, and parents; decrease the discouragement of internet-searching from providers; and guide parents toward reliable and credible internet resources for information and emotional support.

Similar to the studies that highlighted the differences in the knowledge of genomic concepts among young children, adolescents, and parents, several studies highlighted the importance of age in decision-making and the preferences related to genetic information [12,38,39]. Although most adolescents wished to be involved in the decision-making process, parents expressed concerns regarding their child's privacy and capacity to understand genetic information. In general, most participants preferred a shared decision-making model involving the child, parents, and health care practitioners. Several studies exploring decision-making views among patients with chronic illnesses have also reported that a shared decision-making model is preferred by adolescents and their parents [46-48]. In particular, research in the population with JIA has shown that adolescents wish to be a part of the decision-making process and that the parents and providers play a key role in involving children in decision-making and educating them about their disease [57,58].

Research regarding the psychological impact of returning genetic information to children has shown mixed results. A review by Wakefield et al [59] found that serious adverse psychological outcomes such as anxiety, depression, and distress from receiving genetic information were uncommon among children; however, some children experienced interfamilial distress, discrimination, and regret. Research has also shown that receiving genetic information during childhood may allow for early psychological adjustment and the ability to share information with other family members [60]. More research is required to assess the psychological impact and ethical implications of returning genetic information to children, especially for conditions that may not be treatable or modifiable [59]. Therefore, although adolescents prefer to be involved in the decision-making process, a shared decision-making model involving the adolescent, parents, and HCPs may be best suited for decision-making involving the return of genetic information for adolescents with genetic illnesses.

Although only 7% (1/14) of studies explored adolescents' experiences with genetic counseling [16], it presents important

implications for sharing genomic information with young adolescents. Adolescents reported key differences in their experiences talking to a genetic counselor compared with an HCP and felt that learning about genetics, inheritance, and origins of the condition helped them contextualize their condition as a part of their identity and helped them understand their disease. Interestingly, the participants also suggested using videos to describe genomic concepts; a study by Sabatello et al [15] also found that a video format was more effective in increasing self-reported genomic knowledge compared with a pamphlet format. Similar to our findings in the knowledge and decision-making categories, the participants reiterated that age was an important factor in their understanding of genetic information and that high school was a more appropriate period to receive genetic counseling compared with elementary school. However, it is important to consider that 1 study is not representative of the general attitudes that adolescents have toward genetic counseling and more research must be done in the field of genetic counseling and patient education among children and adolescents.

Although the primary purpose of our research was to investigate the knowledge of genomic concepts, in particular, most results from our search returned studies focused on genetic testing, perhaps owing to the novelty of the field of genomic education among pediatric populations. We leveraged a broad search strategy to encompass findings from various fields of genomics health literacy and patient education, which can be leveraged to inform the design of digital health interventions for genomics education. In addition to conducting formative research for the UCAN study, The Centre for Global eHealth Innovation at the University Health Network has developed and conducted usability studies for multiple patient-centric digital applications that aid in the self-management of chronic diseases such as diabetes [61], asthma [62], arthritis [63], prostate cancer [64], and heart failure [65]. The overarching theme among these studies has been to find strategies to engage and educate patients about their chronic illness to facilitate informed decision-making and disease self-management. Thus, the following findings (in no particular order) presented in Textbox 2 could be considered for developing design principles for digital health interventions.

Textbox 2. Key findings for informing the design of patient education digital health interventions.

Key findings

- Providing the patients and parents a high-level overview of genomic concepts relevant to their condition, supplemented by an overview of genetics if applicable to the patient.
- Curating and separating educational content for different age groups (young children, adolescents, and parents) based on their differing capacity to understand genomic information.
- Using visual aids such as illustrations and videos to facilitate engagement.
- Testing the understanding of genomic and genetic concepts without creating the pressure of a test environment, for example, by leveraging quizzes such as the Genetic Knowledge Index before and after providing the patient and parents with educational materials or through a deeper discussion among patients, parents, health care providers or nurses to evaluate the understanding of genomic and genetic concepts.
- Providing a repository of credible and reliable web resources for patients and their parents to seek information relevant to their condition and to seek emotional support in the community.

Most studies in this review were qualitative or mixed methods studies and used methods such as interviews or surveys to gather data from patients, which likely reflects the state of science in pediatric populations and, for most genetic illnesses, remains at the purely exploratory or descriptive level at this time. A limitation of our study is that our understanding of patient education interventions is still limited as only 14% (2/14) of studies leveraged an intervention design to evaluate the change in genetic knowledge of the participants [3,41]. A key limitation of the intervention by Newcomb et al [3] was that the participants were not asked to complete the GKI questionnaire after the test and the understanding of concepts such as *DNA* and *genes* was assessed qualitatively among the participants, which may have led to subjectivity and bias in reporting of posttest results. Another limitation was that the study by Johnson et al [41] investigated learning in parents with a mean age of 37.5 years. Thus, the findings from the study by Johnson et al [41] may not be applicable to inform the design of patient education interventions for children. More research that incorporates intervention methodology to evaluate participants' genomic health literacy levels and attitudes is required to evaluate the best tools for improving genomic literacy in children.

Another limitation of our review is that 29% (4/14) of the included studies investigated the understanding of genetic concepts among populations that do not have any specified disease [12,37-39]. Although it is valuable to gauge the level of genetic knowledge and understanding in the general public, the knowledge and attitudes toward receiving and seeking genomic information may be very different for families in which a family member has a medical condition compared with populations that are not diagnosed with a genetic condition. Moreover, of the 14 studies, 8 (57%) included parent study populations and only 6 (43%) included child and adolescent populations, which may be because of the ethical barriers that currently exist in returning genetic information to children or evaluating their role in genetic decision-making [20,21].

In addition, a key consideration is that we used a broad definition for the terms *genetic* and *genomic* information for the inclusion criteria, which align with the definition of genomic health literacy: "basic knowledge of genetic and genomic concepts and the capacity to obtain, process, understand, and use genomic information for health-related decision-making" [6]. The rationale behind using a broad definition and including study populations with a broad range of genetic and genomic

illnesses was to capture a diverse range of ongoing initiatives that aim to improve genomic health literacy. However, the key findings to curate design principles for digital health interventions are considerations to create digital health interventions for all pediatric populations with genetic conditions and not specific to a singular genetic illness.

Furthermore, the inclusion criteria were limited to studies from the past 12 years to capture lead practices and emerging trends in digital health interventions and genomics. However, a larger sample of relevant studies on patient education and genomic health literacy may have been captured if studies older than 12 years were also included.

Moreover, the ability to cross-culturally compare the knowledge and understanding of genetic concepts was limited because most of the studies were conducted in the United States and only studies published in English were included in this review. This may have also led to bias when reporting results of genomic knowledge among children and their parents as educational systems vary across different countries.

Conclusions

Our review indicates that although pediatric patients and their parents have a positive attitude toward learning genomic information, we still have little knowledge about the genomic health literacy levels among children and adolescents, their capacity to understand genomic concepts, how this information can be presented, and what best practices can be leveraged to design digital health patient education interventions for genomic education. The rise of the personalized and precision medicine approach demands more patient and parent engagement, and it is the medical world's mandate to develop tools that improve patient education on disease knowledge and genomic factors involved. Thus, there is a need for studies that examine the genomic health literacy and modalities to inform the design of digital interventions that aim to educate adolescents and children with pediatric illnesses about genomics.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 91 KB - jmir_v23i12e26684_app1.pdf]

Multimedia Appendix 2
Data extraction table.
[XLSX File (Microsoft Excel File), 21 KB - jmir_v23i12e26684_app2.xlsx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.
[PDF File (Adobe PDF File), 276 KB - jmir_v23i12e26684_app3.pdf]

References

1. Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? J Transl Med 2019 Apr 05;17(1):114 [FREE Full text] [doi: 10.1186/s12967-019-1864-9] [Medline: 30953518]
2. Feero WG, Guttmacher AE. Genomics, personalized medicine, and pediatrics. Acad Pediatr 2014 Jan;14(1):14-22 [FREE Full text] [doi: 10.1016/j.acap.2013.06.008] [Medline: 24369865]
3. Newcomb P, Hudlow R, Heilskov J, Martinez CD, Le H. Conversations with children about DNA and genes using an original children's book. J Pediatr Health Care 2014 Nov;28(6):497-506. [doi: 10.1016/j.pedhc.2014.02.006] [Medline: 24786583]
4. Tse C, Sahhar M, Wallace J, Duncan RE. "There's a whole different way of working with adolescents": interviews with Australian Genetic Counselors about their experiences with adolescent clients. J Genet Couns 2013 Oct 18;22(5):674-684. [doi: 10.1007/s10897-013-9601-7] [Medline: 23686516]
5. Duncan RE, Young M. Tricky teens: are they really tricky or do genetic health professionals simply require more training in adolescent health? Per Med 2013 Aug;10(6):589-600. [doi: 10.2217/pme.13.49] [Medline: 29776191]
6. Hurle B, Citrin T, Jenkins JF, Kaphingst KA, Lamb N, Roseman JE, et al. What does it mean to be genomically literate?: National Human Genome Research Institute Meeting Report. Genet Med 2013 Aug 28;15(8):658-663 [FREE Full text] [doi: 10.1038/gim.2013.14] [Medline: 23448722]
7. McGill B, Wakefield C, Vetsch J, Barlow-Stewart K, Kasparian N, Patenaude A, et al. Children and young people's understanding of inherited conditions and their attitudes towards genetic testing: a systematic review. Clin Genet 2019 Jan 18;95(1):10-22. [doi: 10.1111/cge.13253] [Medline: 29574695]
8. Ulph F, Leong J, Glazebrook C, Townsend E. A qualitative study exploring genetic counsellors' experiences of counselling children. Eur J Hum Genet 2010 Oct 9;18(10):1090-1094 [FREE Full text] [doi: 10.1038/ejhg.2010.86] [Medline: 20531440]

9. Driessnack M. Growing up at the intersection of the genomic era and the information age. J Pediatr Nurs 2009 Jun;24(3):189-193. [doi: 10.1016/j.pedn.2007.09.008] [Medline: 19467431]
10. Mackert M, Rew L, Bonevac D, Champlin S. Older adolescents' perceptions and intentions regarding Do-It-Yourself Genetic Assessment services. J Spec Pediatr Nurs 2012 Apr;17(2):159-167. [doi: 10.1111/j.1744-6155.2012.00329.x] [Medline: 22463476]
11. Hershberger PE, Gallo AM, Molokie R, Thompson AA, Suarez ML, Yao Y, et al. Perception of young adults with sickle cell disease or sickle cell trait about participation in the CHOICES randomized controlled trial. J Adv Nurs 2016 Jun 15;72(6):1430-1440 [FREE Full text] [doi: 10.1111/jan.12702] [Medline: 26073718]
12. McGowan ML, Prows CA, DeJonckheere M, Brinkman WB, Vaughn L, Myers MF. Adolescent and parental attitudes about return of genomic research results: focus group findings regarding decisional preferences. J Empir Res Hum Res Ethics 2018 Oct 28;13(4):371-382 [FREE Full text] [doi: 10.1177/1556264618776613] [Medline: 29806518]
13. Harris ED, Ziniel SI, Amatruda JG, Clinton CM, Savage SK, Taylor PL, et al. The beliefs, motivations, and expectations of parents who have enrolled their children in a genetic biorepository. Genet Med 2012 Mar 26;14(3):330-337 [FREE Full text] [doi: 10.1038/gim.2011.25] [Medline: 22241099]
14. Sapp J, Dong D, Stark C, Ivey L, Hooker G, Biesecker L, et al. Parental attitudes, values, and beliefs toward the return of results from exome sequencing in children. Clin Genet 2014 Feb 20;85(2):120-126 [FREE Full text] [doi: 10.1111/cge.12254] [Medline: 24033230]
15. Sabatello M, Chen Y, Sanderson SC, Chung WK, Appelbaum PS. Increasing genomic literacy among adolescents. Genet Med 2019 Apr 14;21(4):994-1000 [FREE Full text] [doi: 10.1038/s41436-018-0275-2] [Medline: 30214064]
16. Pichini A, Shuman C, Sappleton K, Kaufman M, Chitayat D, Babul-Hirji R. Experience with genetic counseling: the adolescent perspective. J Genet Couns 2016 Jun 17;25(3):583-595. [doi: 10.1007/s10897-015-9912-y] [Medline: 26573304]
17. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 2013 Jul 20;15(7):565-574 [FREE Full text] [doi: 10.1038/gim.2013.73] [Medline: 23788249]
18. Bush LW, Bartoshesky LE, David KL, Wilfond B, Williams JL, Holm IA, ACMG SELI committee. Pediatric clinical exome/genome sequencing and the engagement process: encouraging active conversation with the older child and adolescent: points to consider-a statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med 2018 Jul 22;20(7):692-694. [doi: 10.1038/gim.2018.36] [Medline: 29565417]
19. Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. Genet Med 2008 Jan;10(1):19-32. [doi: 10.1097/gim.0b013e31815f524f]
20. Botkin J, Belmont J, Berg J, Berkman B, Bombard Y, Holm I, et al. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. Am J Hum Genet 2015 Jul 02;97(1):6-21 [FREE Full text] [doi: 10.1016/j.ajhg.2015.05.022] [Medline: 26140447]
21. Braverman G, Shapiro ZE, Bernstein JA. Ethical issues in contemporary clinical genetics. Mayo Clin Proc Innov Qual Outcomes 2018 Jun;2(2):81-90 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2018.03.005] [Medline: 30225437]
22. Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. Lancet 2011 Jun;377(9783):2138-2149. [doi: 10.1016/S0140-6736(11)60244-4]
23. Weiss JE, Luca NJ, Boneparth A, Stinson J. Assessment and management of pain in juvenile idiopathic arthritis. Paediatr Drugs 2014 Dec 21;16(6):473-481. [doi: 10.1007/s40272-014-0094-0] [Medline: 25331986]
24. Barut K, Adrovic A, Şahin S, Kasapçopur Ö. Juvenile idiopathic arthritis. Balkan Med J 2017 Apr 05;34(2):90-101 [FREE Full text] [doi: 10.4274/balkanmedj.2017.0111] [Medline: 28418334]
25. Giancane G, Swart JF, Castagnola E, Groll AH, Horneff G, Huppertz H, Paediatric Rheumatology International Trials Organisation (PRINTO). Opportunistic infections in immunosuppressed patients with juvenile idiopathic arthritis: analysis by the Pharmachild Safety Adjudication Committee. Arthritis Res Ther 2020 Apr 07;22(1):71 [FREE Full text] [doi: 10.1186/s13075-020-02167-2] [Medline: 32264969]
26. Jacobson JL, Pham JT. Juvenile idiopathic arthritis: a focus on pharmacologic management. J Pediatr Health Care 2018 Sep;32(5):515-528. [doi: 10.1016/j.pedhc.2018.02.005] [Medline: 30177013]
27. Ashida S, Goodman M, Pandya C, Koehly L, Lachance C, Stafford J, et al. Age differences in genetic knowledge, health literacy and causal beliefs for health conditions. Public Health Genomics 2011;14(4-5):307-316 [FREE Full text] [doi: 10.1159/000316234] [Medline: 20829577]
28. Lea D, Kaphingst K, Bowen D, Lipkus I, Hadley D. Communicating genetic and genomic information: health literacy and numeracy considerations. Public Health Genomics 2011;14(4-5):279-289 [FREE Full text] [doi: 10.1159/000294191] [Medline: 20407217]
29. Erby LH, Roter D, Larson S, Cho J. The rapid estimate of adult literacy in genetics (REAL-G): a means to assess literacy deficits in the context of genetics. Am J Med Genet A 2008 Jan 15;146A(2):174-181. [doi: 10.1002/ajmg.a.32068] [Medline: 18076116]
30. Meilleur KG, Littleton-Kearney MT. Interventions to improve patient education regarding multifactorial genetic conditions: a systematic review. Am J Med Genet A 2009 Feb 15;149A(4):819-830 [FREE Full text] [doi: 10.1002/ajmg.a.32723] [Medline: 19291763]

31. Bogaard K, Johnson M. Genetics and genomic medicine. Radiol Technol 2009;80(4):318-336. [Medline: 19293120]
32. 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]
33. Barton KS, Wingerson A, Barzilay JR, Tabor HK. "Before Facebook and before social media…we did not know anybody else that had this": parent perspectives on internet and social media use during the pediatric clinical genetic testing process. J Community Genet 2019 Jul 19;10(3):375-383 [FREE Full text] [doi: 10.1007/s12687-018-0400-6] [Medline: 30569339]
34. Gallo AM, Knafl KA, Angst DB. Information management in families who have a child with a genetic condition. J Pediatr Nurs 2009 Jun;24(3):194-204 [FREE Full text] [doi: 10.1016/j.pedn.2008.07.010] [Medline: 19467432]
35. Roche MI, Skinner D. How parents search, interpret, and evaluate genetic information obtained from the internet. J Genet Couns 2009 Apr 21;18(2):119-129 [FREE Full text] [doi: 10.1007/s10897-008-9198-4] [Medline: 18937062]
36. Schaffer R, Kuczynski K, Skinner D. Producing genetic knowledge and citizenship through the Internet: mothers, pediatric genetics, and cybermedicine. Sociol Health Illn 2008 Jan;30(1):145-159 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01042.x] [Medline: 18254838]
37. Lewis C, Loe BS, Sidey-Gibbons C, Patch C, Chitty LS, Sanderson SC. Development of a measure of genome sequencing knowledge for young people: the kids-KOGS. Clin Genet 2019 Nov 30;96(5):411-417 [FREE Full text] [doi: 10.1111/cge.13607] [Medline: 31323115]
38. Myers MF, Martin LJ, Prows CA. Adolescents' and parents' genomic testing decisions: associations with age, race, and sex. J Adolesc Health 2020 Mar;66(3):288-295 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.08.028] [Medline: 31685375]
39. Rew L, Mackert M, Bonevac D. Cool, but is it credible? Adolescents' and parents' approaches to genetic testing. West J Nurs Res 2010 Aug 06;32(5):610-627 [FREE Full text] [doi: 10.1177/0193945909360781] [Medline: 20693337]
40. Khan CM, Moore EG, Leos C, Rini C. Patient hopes for diagnostic genomic sequencing: roles of uncertainty and social status. Eur J Hum Genet 2016 Jun 23;24(6):803-808 [FREE Full text] [doi: 10.1038/ejhg.2015.204] [Medline: 26395557]
41. Johnson L, Sykes AD, Lu Z, Valdez JM, Gattuso J, Gerhardt E, et al. Speaking genomics to parents offered germline testing for cancer predisposition: use of a 2-visit consent model. Cancer 2019 Jul 15;125(14):2455-2464 [FREE Full text] [doi: 10.1002/cncr.32071] [Medline: 30901077]
42. Fitzgerald-Butt SM, Klima J, Kelleher K, Chisolm D, McBride KL. Genetic knowledge and attitudes of parents of children with congenital heart defects. Am J Med Genet A 2014 Dec 24;164A(12):3069-3075 [FREE Full text] [doi: 10.1002/ajmg.a.36763] [Medline: 25256359]
43. Campbell T, Chen C, Chhina H, Chahal R, Cooper A, Elliott A. Support, information, and integration of genetics for children with congenital lower limb deficiencies in British Columbia, Canada. Paediatr Child Health 2019 Sep;24(6):395-401 [FREE Full text] [doi: 10.1093/pch/pxz001] [Medline: 31528111]
44. 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]
45. Furr LA, Kelly SE. The Genetic Knowledge Index: developing a standard measure of genetic knowledge. Genet Test 1999 Jan;3(2):193-199. [doi: 10.1089/gte.1999.3.193] [Medline: 10464667]
46. Dougherty M, Pleasants C, Solow L, Wong A, Zhang H. A comprehensive analysis of high school genetics standards: are states keeping pace with modern genetics? CBE Life Sci Educ 2011 Sep;10(3):318-327 [FREE Full text] [doi: 10.1187/cbe.10-09-0122] [Medline: 21885828]
47. National curriculum in England: science programmes of study. Department for Education. 2020. URL: https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study [accessed 2021-11-26]
48. The Ontario curriculum, grades 11 and 12: science (revised). Ministry of Education. 2008. URL: http://www.edu.gov.on.ca/eng/curriculum/secondary/2009science11_12.pdf [accessed 2021-11-26]
49. van Pelt PA, Drossaert CH, Kruize AA, Huisman J, Dolhain RJ, Wulffraat NM. Use and perceived relevance of health-related Internet sites and online contact with peers among young people with juvenile idiopathic arthritis. Rheumatology (Oxford) 2015 Oct 20;54(10):1833-1841. [doi: 10.1093/rheumatology/kev193] [Medline: 25998449]
50. Stinson JN, Toomey PC, Stevens BJ, Kagan S, Duffy CM, Huber A, et al. Asking the experts: exploring the self-management needs of adolescents with arthritis. Arthritis Rheum 2008 Jan 15;59(1):65-72 [FREE Full text] [doi: 10.1002/art.23244] [Medline: 18163408]
51. Marshall DA, MacDonald KV, Heidenreich S, Hartley T, Bernier FP, Gillespie MK, et al. The value of diagnostic testing for parents of children with rare genetic diseases. Genet Med 2019 Dec 26;21(12):2798-2806. [doi: 10.1038/s41436-019-0583-1] [Medline: 31239560]
52. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001 May 23;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
53. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]

54. Murray E, Lo B, Pollack L, Donelan K, Catania J, White M, et al. The impact of health information on the internet on the physician-patient relationship: patient perceptions. Arch Intern Med 2003 Jul 28;163(14):1727-1734. [doi: 10.1001/archinte.163.14.1727] [Medline: 12885689]
55. Tuffrey C, Finlay F. Use of the internet by parents of paediatric outpatients. Arch Dis Child 2002 Dec;87(6):534-536 [FREE Full text] [doi: 10.1136/adc.87.6.534] [Medline: 12456558]
56. van Dijkhuizen EH, Egert T, Egert Y, Costello W, Schoemaker C, Fernhout M, et al. Patient's experiences with the care for juvenile idiopathic arthritis across Europe. Pediatr Rheumatol Online J 2018 Feb 08;16(1):10 [FREE Full text] [doi: 10.1186/s12969-018-0226-0] [Medline: 29422094]
57. Lipstein EA, Muething KA, Dodds CM, Britto MT. "I'm the one taking it": adolescent participation in chronic disease treatment decisions. J Adolesc Health 2013 Aug;53(2):253-259. [doi: 10.1016/j.jadohealth.2013.02.004] [Medline: 23561895]
58. Lipstein EA, Dodds CM, Lovell DJ, Denson LA, Britto MT. Making decisions about chronic disease treatment: a comparison of parents and their adolescent children. Health Expect 2016 Jun 03;19(3):716-726 [FREE Full text] [doi: 10.1111/hex.12210] [Medline: 24889468]
59. Wakefield CE, Hanlon LV, Tucker KM, Patenaude AF, Signorelli C, McLoone JK, et al. The psychological impact of genetic information on children: a systematic review. Genet Med 2016 Aug 7;18(8):755-762. [doi: 10.1038/gim.2015.181] [Medline: 26741411]
60. Aatre RD, Day SM. Psychological issues in genetic testing for inherited cardiovascular diseases. Circ Cardiovasc Genet 2011 Feb;4(1):81-90. [doi: 10.1161/circgenetics.110.957365]
61. 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] [Medline: 26455762]
62. Morita PP, Yeung MS, Ferrone M, Taite AK, Madeley C, Lavigne AS, 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]
63. Lalloo C, Harris L, Hundert A, Berard R, Cafazzo J, Connelly M, et al. The iCanCope pain self-management application for adolescents with juvenile idiopathic arthritis: a pilot randomized controlled trial. Rheumatology (Oxford) 2021 Jan 05;60(1):196-206. [doi: 10.1093/rheumatology/keaa178] [Medline: 32613229]
64. Pham Q, Cafazzo JA, Feifer A. Adoption, acceptability, and effectiveness of a mobile health app for personalized prostate cancer survivorship care: protocol for a realist case study of the Ned App. JMIR Res Protoc 2017 Oct 12;6(10):e197 [FREE Full text] [doi: 10.2196/resprot.8051] [Medline: 29025699]
65. Seto E, Morita PP, Tomkun J, Lee TM, Ross H, Reid-Haughian C, et al. Implementation of a heart failure telemonitoring system in home care nursing: feasibility study. JMIR Med Inform 2019 Jul 26;7(3):e11722 [FREE Full text] [doi: 10.2196/11722] [Medline: 31350841]

Abbreviations

GKI: Genetic Knowledge Index
HCP: health care provider
JIA: juvenile idiopathic arthritis
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UCAN: Understanding Childhood Arthritis Network

Edited by R Kukafka; submitted 04.01.21; peer-reviewed by WD Dotson, A Banerjee, M Giri; comments to author 10.05.21; revised version received 21.05.21; accepted 10.11.21; published 24.12.21.

Please cite as:
Gupta A, Cafazzo JA, IJzerman MJ, Swart JF, Vastert S, Wulffraat NM, Benseler S, Marshall D, Yeung R, Twilt M
Genomic Health Literacy Interventions in Pediatrics: Scoping Review
J Med Internet Res 2021;23(12):e26684
URL: https://www.jmir.org/2021/12/e26684
doi:10.2196/26684
PMID:34951592

Review

The Role of Recipient Characteristics in Health Video Communication Outcomes: Scoping Review

Daniel Adrian Lungu[1], PhD; Jo Røislien[1], PhD; Siri Wiig[1], PhD; Marie Therese Shortt[1], PhD; Francesca Ferrè[2], PhD; Siv Hilde Berg[1], PhD; Henriette Thune[1], PhD; Kolbjørn Kallesten Brønnick[1], PhD

[1]SHARE – Centre for Resilience in Healthcare, Department of Quality and Health Technology, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

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

Corresponding Author:
Daniel Adrian Lungu, PhD
SHARE – Centre for Resilience in Healthcare, Department of Quality and Health Technology
Faculty of Health Sciences
University of Stavanger
PO Box 8600
Stavanger, 4036
Norway
Phone: 47 51831000
Email: daniel.a.lungu@uis.no

Abstract

Background: The importance of effective communication during public health emergencies has been highlighted by the World Health Organization, and it has published guidelines for effective communication in such situations. With video being a popular medium, video communication has been a growing area of study over the past decades and is increasingly used across different sectors and disciplines, including health. Health-related video communication gained momentum during the SARS-CoV-2 pandemic, and video was among the most frequently used modes of communication worldwide. However, although much research has been done regarding different characteristics of video content (the message) and its delivery (the messenger), there is a lack of knowledge about the role played by the characteristics of the recipients for the creation of effective communication.

Objective: The aim of this review is to identify how health video communication outcomes are shaped by recipient characteristics, as such characteristics might affect the effectiveness of communication. The main research question of the study is as follows: do the characteristics of the recipients of health videos affect the outcomes of the communication?

Methods: A scoping review describing the existing knowledge within the field was conducted. We searched for literature in 3 databases (PubMed, Scopus, and Embase) and defined eligibility criteria based on the relevance to the research question. Recipient characteristics and health video communication outcomes were identified and classified.

Results: Of the 1040 documents initially identified, 128 (12.31%) met the criteria for full-text assessment, and 39 (3.75%) met the inclusion criteria. The included studies reported 56 recipient characteristics and 42 communication outcomes. The reported associations between characteristics and outcomes were identified, and the potential research opportunities were discussed. Contributions were made to theory development by amending the existing framework of the Integrated-Change model, which is an integrated model of motivational and behavioral change.

Conclusions: Although several recipient characteristics and health video communication outcomes were identified, there is a lack of robust empirical evidence on the association between them. Further research is needed to understand how the preceding characteristics of the recipients might affect the various outcomes of health video communication.

(J Med Internet Res 2021;23(12):e30962) doi:10.2196/30962

KEYWORDS

health communication; video communication; communication outcomes; recipient characteristics; recipient factors; health video communication

Introduction

Communication in Public Health Emergencies

Effective communication in public health emergencies is crucial, as people need to not only know but hopefully also understand the health risks they face and what actions they can take to protect themselves, their close ones, and the society from health hazards. The importance of communication in public health emergencies has been highlighted by the World Health Organization in the guidelines for risk communication in public health emergencies policy and practice [1]. The guidelines reflect the complexity of the topic, as several dimensions prove to determine the effectiveness of communication and must therefore be taken into consideration. Dimensions include information accuracy, timeliness and frequency of communication, clarity of language, use of appropriate media channels, building trust with local communities, the use of visuals in combination with—or instead of—text, and the use of new communication channels such as social media. Adding to the complexity, multimedia approaches have been found to be more effective than single media approaches [2].

In recent years, video communication has received increased attention across multiple fields, from education to science, risk, and health communication. Video allows rapid communication, is flexible and able to incorporate empathy, and has good outreach potential. Video communication gained further momentum during the COVID-19 pandemic, and most of the communication aimed at the population was through video. The effectiveness of delivering education through video has been widely investigated even before 2019 [3-7]; however, the COVID-19 pandemic has magnified and speeded up its adoption, and consequently, we can see increased research efforts in the area of effective communication using video [8-14].

Health Communication

Health communication, which is defined as the dissemination and interpretation of health-related messages [15], is a well-established research area, with >300,000 search results in Google Scholar. A total of 4 elements have emerged from this extensive body of literature. The first element is the pervasive effort of theorizing health communication and putting it into practice [16-21]. The second element is the importance of cultural context for the planning and effectiveness of health communication interventions. On the one hand, different cultures require different health communication strategies [22,23]; however, within the same cultural context, different population groups need diversified communications to ensure effectiveness [24-26]. Thus, international literature often mentions the terms *tailored* and *targeted* health communication strategies designed to enhance the relevance of health information to a given audience [27,28]. The third element is empirical evidence concerning the effectiveness of health communication in achieving behavioral change or other public health goals (eg, the eradication of polio and other immunization coverage interventions) [29-32]. The fourth element is the challenge of reshaping and adapting health communication to reflect technological developments and societal changes. Indeed, the web-based environment, social media, advancements in artificial intelligence, and developments of health care (eg, eHealth and telemedicine) pose a challenge to health communication in the information age [33-37]. Furthermore, societies—especially those with universal health care systems—are called to tackle the health inequalities arising from technological advancements. Health communication must consider the digital divide and the different digital health literacies of the population to be effective [38]. This can be extremely relevant during pandemics, as, in the case of the COVID-19 pandemic, the groups affected by the digital divide and generally having less digital health literacy were the most affected [39-41].

Although much research has been conducted on video communication itself, there is less evidence regarding whether the outcomes (eg, knowledge, attitudes, compliance, and behavior) of health video communication are affected by the various characteristics (eg, sociodemographic, personality and values, and environmental factors) of the *recipients* of this communication. Health video communication refers to communicating health information through video format regardless of style (eg, animation, text over image, interviews, and voice over image). Health communication is multifaceted and variably affected by the messenger, message attributes, and recipient characteristics (eg, sociodemographic, personality and values, and environmental factors). This study aims to address this gap by identifying the relevant recipient characteristics and health video communication outcomes reported in the literature and the relationship between them. In this review, the concept of health video communication does not refer solely to public health communication videos but rather to health videos in general, including instructional videos for patient training and education, videos from public health prevention campaigns, and in-hospital informational videos.

A Health Communication Framework

For this study, we chose the Integrated-Change (I-Change) model (Figure 1) developed by Hein de Vries [42] as the study's theoretical framework. As the driving research question is to identify recipient characteristics, outcomes of health video communication, and any relationship between them, the I-Change model fits the research question and the methodological approach of this study. Moreover, the categorization presented in the model is useful for discussing results and understanding the current state of knowledge and research practice. The I-Change model is an integrated model of motivational and behavioral change that combines elements from the theory of planned behavior [43], the social cognitive theory [44], the Transtheoretical Model of Health Behavior Change [45], the Health Belief Model [46,47], and the goal-setting theory [48,49]. The integration of these elements into a single health communication framework has made the adoption of the I-Change model appropriate as a basis for this study. The model states that both covert (hidden and not observable) and overt (visible and observable) behavior is determined by a person's motivation or intention. In turn, motivation depends on attitudes, social influences, and self-efficacy. Attitudes are defined as the perceived cognitive and emotional advantages and disadvantages of a person's behavior. A person's social influences comprise social modeling (the perception of others having this behavior), social norms

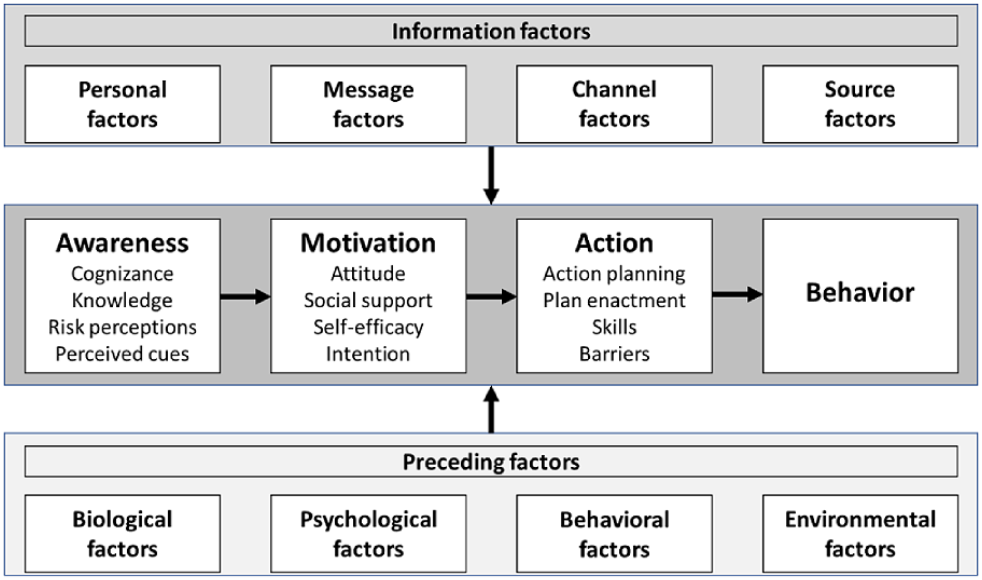
(the norms that people have regarding this behavior), and social support received from others when performing the behavior. Self-efficacy refers to the personal judgment of how well one can execute the courses of action to deal with prospective situations [44].

The model assumes that the communication outcomes (awareness, motivation, action, and behavior) depend on 2 determinants: information factors and preceding factors.

The information factors (Figure 1) have long been debated in international literature, and the importance of the source, channel, message, and messenger's personal factors in influencing the outcomes of communication has been demonstrated in a wide range of disciplines, including, but not limited to, health, education, marketing, and risk communication [23,50-54].

The preceding factors (Figure 1) are predisposing factors of the recipients of the communication and comprise four categories: biological factors (eg, gender or sex and ethnicity), psychological factors (eg, personality, depression, and anxiety), behavioral factors (eg, lifestyle and adherence to recommendations), and environmental factors (eg, public health policies and availability or lack of resources at the community level). We are aware that gender is not a biological factor and is distinct from sex, but for the purpose of this study, we chose not to further elaborate on the difference between the 2 concepts. The terminology introduced by the I-Change model was adopted, and throughout the article, the terms *preceding factors* and *characteristics* of recipients are used interchangeably. This study focuses on the influence of preceding factors on health video communication outcomes.

Figure 1. The Integrated-Change model.



Methods

Overview

The literature on recipient characteristics is heterogenous with regard to subject groups, methods, research questions, and disciplines. Thus, a scoping review approach was chosen to identify the nature and degree of evidence available in the international literature. Our review followed the five stages according to the methodological framework developed by Arksey and O'Malley [55], which we chose for its completeness: (1) identification of the research questions, (2) identification of relevant studies, (3) selection of studies, (4) data extraction and management, and (5) summary and analysis of results.

Research Questions

This scoping review aimed to explore the role that recipient characteristics (biological, psychological, behavioral, and environmental factors) play in the outcomes of health video communication by (1) identifying which recipient characteristics and outcomes of health video communication were reported in peer-reviewed literature, (2) adapting the I-Change model to health video communication [42], (3) providing a categorization of both recipient characteristics and outcomes into the adapted model, and (4) investigating the relationships between recipient characteristics and health video communication outcomes reported in the international literature. This process will contribute to the existing knowledge on health communication by providing a comprehensive framework for health video communication effectiveness studies. The research questions driving this study were as follows:

1. What are the characteristics of recipients that might influence the outcomes of health video communication?
2. What are the outcomes of health video communication?
3. What relationships exist between the recipient characteristics and the outcomes of health video communication?

Literature Identification

We searched three main databases for public health, social sciences, and biomedical studies on health video communication: PubMed, Scopus, and Embase. Additional

literature was identified using the snowball method [56,57]. The relatively large number of papers identified in the initial search led us to limit the search to peer-reviewed literature. The search string was designed by relying on the population-concept-context framework, as suggested by the scoping reviews chapter of the Joanna Briggs Institute Manual for Evidence Synthesis [58]. The framework comprises defining a string for each component and subsequently combining them into the final search string. The string for the population component was (*people OR patient* OR recipient* OR receiver* OR viewer* OR person*).

The string referring to the concept was (*[past OR previous OR prior] AND [knowledge OR experience]) OR ([person* OR individual*] AND [characteristic* OR propert* OR value*]) AND (outcome* OR attitude* OR behavi* OR accept* OR learn* OR react* OR respon**).

The string for the context component was (*communic* AND video AND health*).

Data from the 3 databases were extracted on December 12, 2020. The search was limited to results published in English only, whereas no restriction was applied to the publication year. The complete search strategy was validated by 2 of the coauthors and is available in Multimedia Appendix 1.

Selection of Studies

The first step of the selection process comprised removing duplicates (approximately one-third) to confirm the completeness of the search string and strategy. In addition to incomplete or unavailable studies, studies were excluded based on title and abstract assessment if (1) they did not focus on recipient characteristics or on specific populations (eg, people who are visually impaired or deaf), (2) they did not focus specifically on health video communication (video used for other purposes, eg, video as a recording tool, referred to video games, video teleconference, video for the education of medical students, video calls, video physician–patient communication, or video simulation), or (3) they did not use video as the main communication method. Studies were included if they met both of the following criteria: (1) they focused on health video communication, and (2) they took into account at least one recipient characteristic. No inclusion or exclusion criteria were set for the design of the identified studies. Studies included in this phase went through full-text assessment, and the final inclusion depended on their coherence with respect to the 2 inclusion criteria mentioned above. The criteria and the results of each stage were made available to all the coauthors, and their feedback was used to solve potential inconsistencies among the scope of the study, the inclusion and exclusion criteria, and the selected articles.

Data Extraction

Microsoft Excel was used to create data extraction forms. For title and abstract assessment, the information stored were the digital object identifier, inclusion or exclusion status, and reason for exclusion. In the second stage of full-text assessment, the information about the publication year, journal, authors, title, document type, country, research domain, recipient characteristics, and reported outcomes were added. All the selected papers were saved using Mendeley, and the library was shared among all the coauthors.

Data Analysis

The extracted data were imported into the NVivo (version 12 Pro; QSR International). The relevant topics, both regarding the recipient characteristics and communication outcomes, were coded. A nested coding methodology [59] was used to organize the information in layers. The relationships between nodes were also coded to keep track of the associations between recipient characteristics and communication outcomes reported in the literature. The found relationships are reported in the description of outcomes in the *Results* section.

After the completion of the coding process, we moved on to define a conceptual framework, starting with the I-Change model and the way it highlights how the 4 dimensions of health communication outcome (awareness, motivation, action, and behavior) are influenced by information factors (eg, message, channel, and source) and preceding factors (biological, psychological, behavioral, and environmental factors). The existence of prior research allowed us to conduct a directed content analysis aimed at validating or conceptually extending the existing theoretical framework, as suggested by Hsieh and Shannon [60]. The content analysis, and especially the identification of key concepts or variables as initial coding categories, was guided by a structured and deductive approach according to the I-Change model [60-62].

The need to adapt the original model emerged at the first attempt to allocate the coded nodes to model categories, as some nodes lacked a fitting category in the original I-Change model. A similar approach of using content analysis to conduct knowledge building and theory development was described by Finfgeld-Connett [63]. The model was refined by an initial discussion between 2 coauthors (DAL and KKB) and then validated by the whole research group.

The final stage of data analysis was to assign nodes to the categories of the refined model. This process was conducted independently by 2 of the coauthors (KKB and DAL), and the divergencies were solved through discussion. In case the discussion did not lead to consensus, a third coauthor (SHB) was involved to have a majority. Finally, the relationships between the nodes reported in the literature were charted.

Results

Search Findings

The database search strategy yielded 1040 records (331/1030, 32.13% from Scopus; 392/1030, 38.05% from PubMed; and 307/1030, 29.81% from the Embase). Approximately 34.66% (357/1030) were duplicates and thus removed, and the title and abstract assessment was performed for the remaining 65.34% (673/1030) of articles. Of the 673 articles, 555 (82.5%) were removed according to the eligibility criteria. A total of 10 additional documents were identified through cross-checking the references. In total, 128 documents were assessed for full-text. Of the 128 papers, 39 (30.5%) met the inclusion criteria and were included in the content analysis. The PRISMA (Preferred Reporting Items for Systematic Reviews and

Meta-Analyses) chart of the selection process is presented in Figure 2. The characteristics of the selected studies are available in Multimedia Appendix 2 [64-100].

All the included studies were published between 2000 and 2020. There is an increasing worldwide publication trend over the past years. As highlighted in Figure 3, the publication trend for the articles included in this study is in line with the international trend, which displays the novelty and growing interest in this research topic.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart of the selection process.

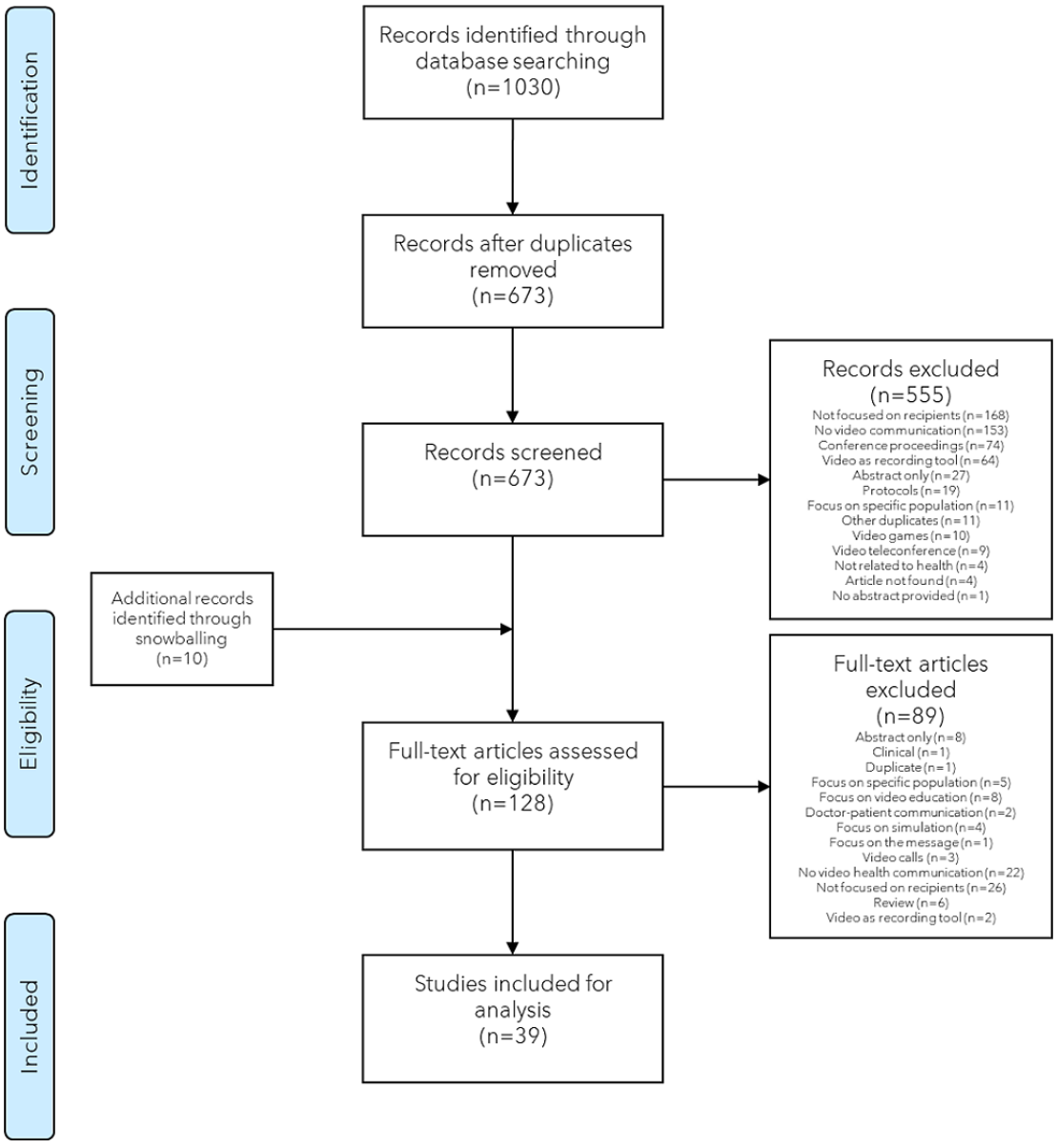
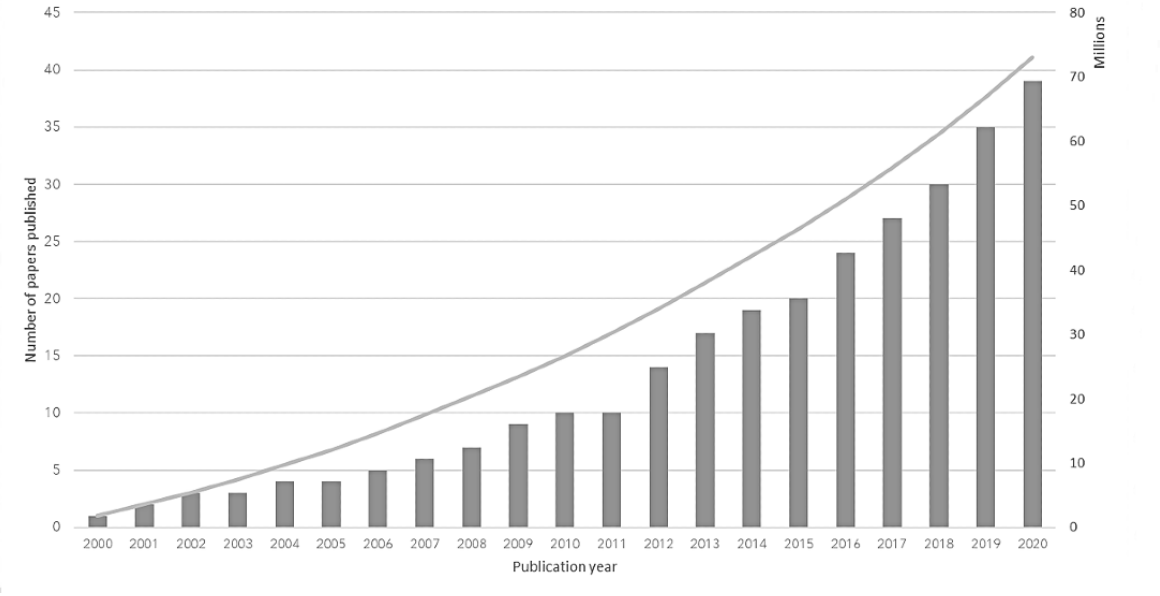


Figure 3. Cumulative number of papers published (worldwide and included in this study). Note that the x- and y-axes have different scales.



Recipient Characteristics

The included studies reported 56 different recipient characteristics. Some characteristics were frequently reported, with age (reported in 33/39, 85% of analyzed articles), ethnicity or race (26/39, 67%), gender or sex (25/39, 64%), and education (24/39, 62%) being the most investigated, whereas others such as sexual orientation, social network, emotional factors, or decisional control preferences being reported only by 3% (1/39) of studies each. Some characteristics were reported by a fair number of studies, such as income or socioeconomic status (SES; 13/39, 33%), beliefs and attitudes (12/39, 31%), level of knowledge (11/39, 28%), disease severity (10/39, 26%), previous experience (10/39, 26%), and health literacy (7/39, 18%). A complete overview of all the characteristics and the number of reporting documents is available in Multimedia Appendix 3, Table S1.

Preceding Factors

Overview

The analysis identified 56 different recipient characteristics. Of the 56 characteristics, 48 (86%) could be assigned to a corresponding category in the I-Change model. However, some factors could not be naturally fitted in the I-Change model and were assigned to a new category termed *Knowledge factors*. Refer the *Additional Category of Knowledge Factors* section for further discussion into this.

Biological Factors

The category of biological factors encompassed 14% (8/56) of the recipient characteristics, namely age, comorbidities, ethnicity or race, gender or sex, self-reported health status, disease severity, current symptoms, and clinical factors. Most of these characteristics were frequently reported by the analyzed studies, with age, ethnicity or race, and gender or sex being the most investigated by international literature. Clinical factors, current symptoms, and comorbidities were rarely investigated as preceding factors that influence health video communication outcomes.

Psychological Factors

Almost half of the identified factors (27/56, 48%) were attributed to the category of psychological factors such as anxiety, depression, decisional control preferences, influential personal factors, information priorities, preferences, values, expectations, psychological distress, beliefs, attitudes, emotional factors, ability to process information, motivation, perception of treatment efficacy, treatment concerns, self-efficacy, trust in information sources, confidence, empathy, trust in health care, health locus of control, regulatory focus, confidence, hesitancy, sexual orientation, hope, and personal relevance. Furthermore, 13% (7/56) factors that were attributed to other categories could also have been included in this category, as they also have psychological implications but were a better fit into other categories. For example, comorbidities, self-reported health, disease severity, current symptoms, and clinical factors were assigned to the biological factors category; risk estimation was attributed to the knowledge factors category; and social norms were assigned to the environmental factors category.

Behavioral Factors

Only 4% (2/56) of the recipient characteristics were ascribed to the category of behavioral factors: involvement and self-reported adherence.

Environmental Factors

The environmental factors category encompassed 20% (11/56) of the identified characteristics. The most recurrent ones were income and SES (reported by 13/39, 33% of the studies), marital status (7/39, 18%), and type of health insurance (6/39, 15%). The remaining factors, generally seldom reported, were the level of information and support that people can rely on, employment, financial comfort, living with someone, social network, social norms, culture, and geographical location.

Additional Category of Knowledge Factors

The added category of knowledge factors contained 14% (8/56) of the factors. Similar to the biological factors, a significant degree of within-group variation was observed for these factors.

Of the 39 included studies, the education level, knowledge, and previous experience were reported by 24 (62%), 11 (28%), and 10 (26%) studies, respectively. On the other hand, information, perceived knowledge, and health history were reported only by 3% (1/39) of the studies analyzed. The category also included 2 relevant factors that received an average amount of coverage in the international literature: health literacy and risk estimation and perception.

Health Video Communication Outcomes

Overview

The included studies reported 42 different health video communication outcomes that covered a wide range of topics. The most frequently reported outcomes were knowledge, including an increase in knowledge and knowledge transfer (15/39, 38%), attitudes (9/39, 23%), behavior (9/39, 23%), and intentions (8/39, 21%). The least frequently included outcomes were awareness, compliance, activation, and attention, reported by only 3% (1/39) of studies each. Some of the outcomes that were reported by a fair number of studies included acceptance (7/39, 18%), beliefs (5/39, 13%), usefulness of the communication (4/39, 10%), and choice of treatment (4/39, 10%). A complete overview of all the health video communication outcomes is available in Multimedia Appendix 3, Table S2.

The 42 health video communication outcomes reported by the included studies were attributed to the categories of the I-Change model as follows: 28 (66%) outcomes were attributed to awareness and motivation (14/42, 33% of outcomes to each), 5 (12%) outcomes were attributed to a new category termed emotions, and the remaining 9 (21%) outcomes were attributed to the category of action that encompassed the original categories of action and behavior. These decisions will be further discussed in *The Revised I-Change Model* section.

Awareness

The most frequently reported outcomes belonged to the awareness category, which encompassed 33% (14/42) of the identified outcomes. Most of the included studies focused on elements of the awareness category as the main outcome of health video communication. The most frequently reported outcome was knowledge, which was explored in 38% (15/39) of the studies. The other included outcomes were expectations, usefulness, recall of information, uncertainty, comprehension, beliefs, perceived benefit, information, information processing, information seeking, perceived risk, perceived prevalence, and awareness.

Most of the reported evidence on awareness was focused on beliefs, knowledge gain, risk and prevalence perception, recall of information, and usefulness of communication. Syrjala et al [64] studied patient training in cancer pain management using integrated print and video materials and found that participants of color were much more likely to report the perception of severe pain than White participants (odds ratio [OR] 18.4, 95% CI 2.1-56.3; *P*=.01). In a study that assessed the effectiveness of smoking cessation communication on YouTube, Romer et al [65] found that ethnicity and current behavior affect beliefs: Hispanic participants were found to have stronger beliefs regarding the mortality effects of smoking, whereas current tobacco use was negatively related to these beliefs. Pretest beliefs, in turn, affected information processing, which is a relevant outcome in health communication [101]. In clinical trials video communication, Curbow et al [66] have found that higher levels of information processing were associated with more negative pretest beliefs about clinical trials. Recall of information was found to be correlated with knowledge factors, namely health literacy and level of education. In an oncology communication setting, Visser et al [67] showed that poor functional health literacy was a predictor for poor information recall when using a standard communication condition without emotional involvement. When the standard was replaced with emotion-oriented communication, health literacy did not have any statistically significant effect. In the context of maternal and neonatal health clips, results showed that the higher the level of education, the greater the information recall [68].

The generation and transfer of knowledge is one of the main health video communication outcomes investigated in international literature and reported by the studies included in this review [69-73]. Evidence regarding the impact of preceding factors on knowledge is mixed. McKenzie et al [74] reported no significant effect of health literacy and education on prepost maternal safety knowledge after exposure to injury prevention video recommendations. Similarly, Bekalu et al [75] investigated the effect of age, gender, education, race or ethnicity, and income on increase in the knowledge of pandemic influenza after video exposure and found no significant effect [75]. A similar result was obtained by Curbow et al [66] in the context of oncology clinical trial communication and by Phelan et al [76] in the context of back surgery, where a randomized trial showed no significant effect of preceding factors on knowledge increase [76]. Grindel et al [77] reported both a nonsignificant impact of preceding factors on breast cancer knowledge after video exposure and a significant drop in knowledge scores after 1 year, meaning that 1-time communication generally leads to increased knowledge in the short run; however, it may not be enough to sustain the knowledge gain in the long term. An opposite finding was reported by Hickey et al [78], who claimed to have achieved long-term (2-month follow-up) knowledge gain after a single viewing of a breast cancer video.

When analyzing the perceived smoking prevalence as an outcome of peer smoking cessation communication, Romer et al [65] found that Hispanic ethnicity (b=10.07; *P*=.02), female gender (b=2.67; *P*=.01), increasing age (b=1.55; *P*=.01), and current tobacco use (b=3.16; *P*<.001) were positively related to perceived smoking prevalence.

The final communication outcome belonging to the awareness category is perceived usefulness, and some studies have reported it to be relevant in the context of educational videos for patients with advanced gastrointestinal cancers [70] and in the context of urologic oncology of patients with localized prostate cancer [79]. However, the article by Albert et al [80] is the only study that investigated the relationship between preceding factors and perceived usefulness. Patients with higher self-confidence in using health devices and less severe conditions (*P*=.02) and patients of color (*P*=.03) perceived higher usefulness of telemonitoring devices. Higher perception of ease of use was

affected by higher health literacy (*P*=.01), previous and current use history (*P*=.01), higher education level (*P*=.03), and married or cohabitating status (*P*=.02).

Motivation

The motivation category encompassed 33% (14/42) of the identified outcomes. Attitudes and intentions were the most investigated outcomes, being explored by 9 (23%) and 8 (21%) of the 39 analyzed studies, respectively. Acceptance, perceived self-confidence, perceived response efficacy, treatment preference, satisfaction, compliance, engagement, confidence, self-efficacy, and relevance are the remaining outcomes, and all of them were reported by less than 10% (4/42) of the selected studies.

More positive attitudes after exposure to video communication have been reported by the included studies in the context of outpatient surgical care and breast, prostate, and colorectal cancer [77,81,82]. Only Engler et al [81] reported that, in the German context, there were differences in the percentage of patients with positive attitudes toward health information on the web, with 92% of patients with colorectal cancer compared with 79% of patients with breast cancer and 53% of patients with prostate cancer.

The topic of confidence after exposure to health videos has been investigated in the United States regarding vaccinations, which is a topic that has been increasingly debated in the context of the COVID-19 pandemic. Nowak et al [83] found overall moderate confidence in the influenza vaccine (mean 2.78 on a scale of 1-5 scale, with *1* corresponding to *not at all confident*), with no differences between communication modalities (video, virtual reality, and electronic pamphlet story), although the influence of preceding factors was not investigated. When analyzing childhood vaccine-related confidence, Mendel-Van Alstyne et al [84] found that familiarity with the vaccine and high SES are associated with high confidence, whereas lacking information about issues that people were concerned about (eg, vaccine ingredients), uncertainty of the interaction of vaccines with children's immune systems, and negative beliefs (eg, that vaccines can cause illnesses) led to low confidence.

Literature from health communication has investigated the effect of individual characteristics on intentions, focusing mainly on biological factors [102]. Results showed that high intentions toward activities that improve health are associated with being a woman [103-105], increasing age [106,107], and being White [108,109]. In the context of health video communication, all the included studies that focused on intentions either did not investigate the effect of preceding factors on them or reported a nonsignificant relationship [72,74,83,85,86].

Emotions

The added category of emotions received fair attention in previous studies, with 12% (5/42) of the identified outcomes belonging to the emotions category. The identified outcomes were emotional response, decisional conflict, decision quality, decision satisfaction, and reactions. Moreover, each emotional outcome was reported by a very limited number of studies, which highlighted the limited amount of focus these factors have attracted and a potential gap in the literature.

There is a general lack of evidence about how emotional factors affect communication outcomes. Although there is some evidence regarding emotional reactions and trust, nothing was reported regarding the acceptance of health messages or the effects on the decision-making process (eg, decision satisfaction and decisional conflict) of recipients in the studies that were eligible for inclusion in this review.

The emotional response to health video communication has been reported to be significantly influenced by the age and gender of the recipients. A study by Prieto-Pinto et al [68] found, by using pupillary dilatation measures, which correlates with autonomous nervous system arousal, that men tend to have a greater emotional reaction after being exposed to maternal and neonatal health video clips. In a study conducted in China that aimed to test video versus virtual reality health communications, Liu et al [87] found that the young and older population differ in positive emotions after being exposed to communication: the young individuals show a greater emotional response to virtual reality, whereas the older individuals show a greater emotional response to health video communication. Of the 39 included studies, trust, which is a key outcome of health communication, was rarely investigated, with only 3 (8%) of the papers included in this review reporting on trust and only 2 (5%) providing empirical evidence regarding the impact of prior mistrust, age, and education on trust. By using a multilevel analysis model in the context of oncology communication, Hillen et al [88] reported a positive effect of age (*P*<.001) and a negative effect of education (*P*=.04) on trust. By applying structural equation modeling to investigate the reactions to survivor of breast cancer stories, McQueen and Kreuter [89] found that the level of medical mistrust at baseline affects the evaluation of the communication video (β=−.29; *P*<.001). The lack of further empirical evidence on trust highlights a potential gap in the literature that future research should address.

Action

The action category, which was obtained by merging the categories of action and behavior of the original I-Change model, encompassed 21% (9/42) of the identified health video communication outcomes. Outcomes belonging to this category received little attention in the literature, with all of them reported by less than 10% (4/42) of the included studies. Capabilities, time to treatment, quality of communication, participation, activation, spreading the message, and behavior complete the list of outcomes included in the action category.

Adherence to treatment is one of the main action outcomes of health video communication. Although adherence was often reported by the studies included in this review, there is little evidence of how the preceding factors affect it. In a study that investigated the risk of nonadherence to antiplatelet medication at the time of coronary stent placement, Palacio et al [90] reported no significant effects of gender, race or ethnicity, health literacy, income, education, and access to care on adherence. Instead, adherence was statistically higher for patients with a spouse or a domestic partner (*P*=.05), with a low Charlson comorbidity score (*P*=.03) and high English proficiency (*P*=.05). Low adherence was observed among those with a positive

screening for depression and among patients with moderate to severe depressive symptoms (*P*=.01).

In the context of maternal vaccine information communication in the United States, Dudley et al [86] investigated the factors associated with referring close contacts after exposure. The likelihood of referring contacts increased for women who intended to receive a maternal influenza vaccine (OR 1.37, 95% CI 1.04-1.81), whereas it decreased for those who were unsure about their infant vaccine intentions (OR 0.47, 95% CI 0.27-0.83). Participants were more likely to refer contacts if they were confident in the safety (OR 1.64, 95% CI 1.13-2.38) and efficacy of the maternal vaccine (OR 1.9, 95% CI 1.21-2.98), had higher perceived susceptibility to (OR 1.62, 95% CI 1.1-2.4) and severity of (OR 2.19, 95% CI 1.28-3.73) influenza during pregnancy, and had trust in the maternal vaccine information from academic institutions (OR 1.56, 95% CI 1.09-2.25) and the infant vaccine information from the Center for Disease Control (OR 1.44, 95% CI 1.02-2.05) and academic institutions (OR 1.85, 95% CI 1.27-2.71). The associations between the likelihood of referring contacts and ethnicity, education, state, or having prior children were not statistically significant.

None of the included studies reported any associations between the preceding factors and the other action outcomes such as activation, behavior, participation, time to treatment, or treatment choice.

Almost one-third (12/39, 31%) of the papers included in this analysis focused on both preceding factors and health video communication outcomes but did not investigate the relationship between them. Information about the preceding factors was collected and used mainly for descriptive purposes rather than explanatory variables [74,91-100,110].

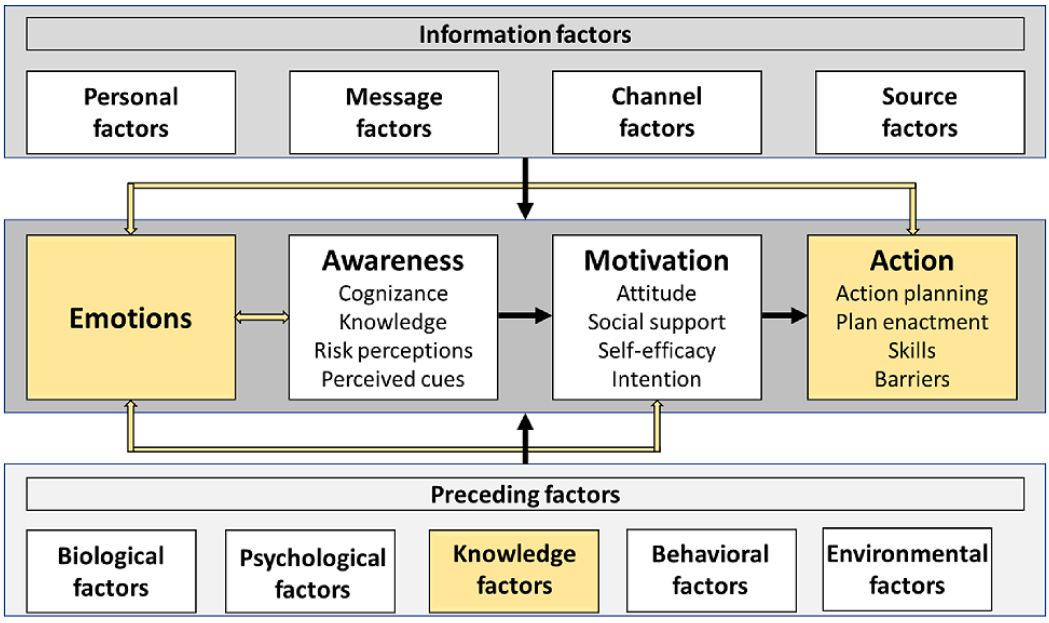
It is noteworthy to observe how the action category received relatively low attention from the included studies, as health videos are often designed and used to drive behavior and behavioral change (eg, smoking cessation, vaccination decisions, and weight loss). Indeed, a tendency to focus on awareness and motivation rather than on the actual behavioral outcome of the recipients emerged from the included studies.

The Revised I-Change Model

When analyzing the included studies—coding the content and identifying relevant nodes regarding both recipient characteristics and communication outcomes—some gaps emerged in the I-Change model when applied to health video communication. Thus, we revised the I-Change model to adapt it to health video communication and obtained an adapted version (Figure 4) that fits the purpose and comprises three main novelties with respect to the original version:

- The inclusion of a fifth category termed knowledge factors within the preceding factors
- The merging of the action and behavior outcome categories into a single category termed action
- The inclusion of an emotions category among the outcomes

Figure 4. The revised Integrated-Change model.



The category of knowledge factors was included as 1 of the preceding factors categories as many included studies reported multiple characteristics associated with knowledge that did not meaningfully belong to biological, psychological, behavioral, or environmental factors. Examples of such characteristics include the recipients' level of knowledge, health literacy, level of information, ability to process information, and information priorities.

The original model foresees action and behavior as distinct communication outcomes, whereas our analysis demonstrated an overlap between the 2 categories using the data from the 39 studies included in this review. Thus, we decided to merge them into a single category termed as action.

The emotional dimension has been shown to play an important role in multiple outcomes of communication: risk perception, judgment, and decision-making [111]. Emotions can both provide significant inputs to judgments and decision-making and can fundamentally change the process of judging and deciding, especially when knowledge about the events is not easily remembered or expressed [112]. Moreover, not only are emotions influenced by preceding and information factors but are also in a mutual relationship with the other outcome categories, as highlighted by the 2-way arrows in Figure 4. This represents an important amendment to the original model, as it considers the feedback effect that awareness, motivation, and action have on emotions.

Discussion

Principal Findings

The included studies provide an extensive overview of the recipients' preceding characteristics that might affect a wide range of health video communication outcomes and present evidence on the relationship between them. The trend observed in Figure 3 displays the novelty and growing interest in this research topic.

Although subsequent analysis revealed that many of the preceding factors belong to the psychological category, only a limited number of papers reported evidence of their impact on outcomes, inviting further research in the area. Biological factors were often reported by the analyzed studies.

Although knowledge and environmental factors received fair attention in international literature, systematic and robust evidence of their effect on communication outcomes is scarce. Moreover, the recipients' first language and their ability to speak and understand the content of videos were surprisingly not identified as relevant factors. This is because the selected studies included only people who were able to speak the language presented in the communication. Further research investigating the magnitude of the impact of linguistic skills on outcomes would positively contribute to the effectiveness of health communication.

The category of behavioral factors appears to be underinvestigated, with only 2 factors identified in all 39 studies and no evidence available on the relationship with outcomes. Although the awareness and motivation categories included the most communication outcomes, those belonging to the action and emotions categories were less reported by the included studies, and there is very little evidence of how they are affected by the preceding factors. Communication and health research would benefit from further efforts focusing on the action category, as it currently represents less than 10% (4/42) of the outcomes identified in our study, with most studies preferring to focus on awareness and motivation outcomes. This number appears low with respect to the general aim of most health communication interventions aimed at changing behavior.

The scarcity of evidence regarding emotions is particularly relevant for effective health communication. Feelings and moods motivate people to reproduce those feelings and moods, whereas, in risk communication, the topic communicated (eg, earthquakes and other natural disasters) usually evokes negative feelings and moods, motivating people to act to avoid them. Emotions are states that are not under voluntary control but are shaped and learned associatively through experiences while being partly innate at birth. Emotions affect attention, memory, motivation, and action [113]; therefore, we included them as a category of preceding factors in the adapted version of the I-Change model.

Our contribution to theory development followed the research approach, in which the theory is understood as emerging from data [114,115]. Indeed, the empirical approach comprises the development of a new theory by relying on empirical observations followed by careful analysis and verification of hypotheses. We started with an existing theoretical framework; however, the empirical data highlighted the misfit of the I-Change model to the topic of health video communication. Therefore, through data analysis, we revised the original model and adapted it to the specific characteristics and needs of health video communication.

As we decided to conduct a scoping review, we are aware of the common limitations of this approach [116]. Information was gathered from a wide range of study designs and methodologies, and the quality of evidence was not formally evaluated. As for the results, it does not provide a synthesized answer to a specific question but rather an overview of the available evidence in the literature.

Further understanding of the gaps presented in this review could have a great impact on the effectiveness of public health emergency communication strategies, as in these contexts, the psychological and behavioral factors of people are key, and emotions are able to significantly affect their decisions and behavior.

Conclusions

Although some evidence of associations between recipients' preceding factors and health video communication has been reported in the literature, our analysis revealed a significant gap in the literature, with many health video communication outcomes and factors not yet explored. This scoping review of the available evidence demonstrated a potential research gap, especially concerning the emotional outcomes of communication and the behavioral and psychological preceding factors of recipients. The review showed that, currently, only a partial picture of the role of recipient characteristics on the outcomes of health video communication is available. Moreover, this study also contributed to the theoretical development of the I-Change model to adapt it to the needs of health video communication.

Acknowledgments

The COVID Communication: Fighting a Pandemic through Translating Science project has received funding from the Trond Mohn Foundation, under grant agreement number TMS2020TMT10 and the University of Stavanger.

Authors' Contributions

This study was conducted by an interdisciplinary team of researchers. DAL and KKB conceived the study, discussed the methodological approach, and analyzed the data. DAL drafted the initial manuscript. All the authors (JR, FF, HT, SW, SHB, and MTS) provided critical feedback on the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete search strategy and results.
[DOCX File , 13 KB - jmir_v23i12e30962_app1.docx]

Multimedia Appendix 2
Overview of included studies.
[DOCX File , 32 KB - jmir_v23i12e30962_app2.docx]

Multimedia Appendix 3
Overview of characteristics and outcomes.
[DOCX File , 20 KB - jmir_v23i12e30962_app3.docx]

References

1. World Health Organization. Communicating risk in public health emergencies: a WHO guideline for emergency risk communication (ERC) policy and practice. Geneva: World Health Organization; 2017.
2. Fitzpatrick-Lewis D, Yost J, Ciliska D, Krishnaratne S. Communication about environmental health risks: a systematic review. Environ Health 2010 Nov 01;9:67 [FREE Full text] [doi: 10.1186/1476-069X-9-67] [Medline: 21040529]
3. Tarasuik J, Galligan R, Kaufman J. Seeing is believing but is hearing? Comparing audio and video communication for young children. Front Psychol 2013;4:64 [FREE Full text] [doi: 10.3389/fpsyg.2013.00064] [Medline: 23450846]
4. Themelis C. Synchronous video communication for distance education: the educators' perspective. Open Praxis 2014;6(3):245-256 [FREE Full text] [doi: 10.5944/openpraxis.6.3.128/]
5. West R, Jay J, Armstrong M, Borup J. "Picturing Them Right in Front of Me": guidelines for implementing video communication in online and blended learning. TechTrends 2017 Jul 1;61(5):461-469. [doi: 10.1007/s11528-017-0208-y]
6. Canto S, Ondarra K. Language learning effects through the integration of synchronous online communication: the case of video communication and second life. Lang Learn High Educ De Gruyter Mouton 2017;7(1):21-53. [doi: 10.1515/cercles-2017-0004]
7. Seckman C. Impact of interactive video communication versus text-based feedback on teaching, social, and cognitive presence in online learning communities. Nurse Educ 2018;43(1):18-22. [doi: 10.1097/NNE.0000000000000448] [Medline: 28858951]
8. Moreno-Guerrero A, Rodríguez-Jiménez C, Gómez-García G, Ramos Navas-Parejo M. Educational innovation in higher education: use of role playing and educational video in future teachers' training. Sustainability 2020 Mar 24;12(6):2558. [doi: 10.3390/su12062558]
9. Dattilo J, Lorek A, Sliwinski M, Chen S, Hill N. An examination of video assisted leisure education in middle-aged and older adults. J Leisure Res 2020 Jun 04;52(2):227-246. [doi: 10.1080/00222216.2020.1767005]
10. Fassbender WJ. The potential for (more-than-)representational video in education research. Int J Res Method Edu 2020 Jun 08;44(3):241-256. [doi: 10.1080/1743727x.2020.1772743]
11. Chen HM, Thomas M. Effects of lecture video styles on engagement and learning. Education Tech Research Dev 2020 Mar 12;68(5):2147-2164. [doi: 10.1007/s11423-020-09757-6]
12. Keswani RN, Sethi A, Repici A, Messmann H, Chiu PW. How to maximize trainee education during the coronavirus disease-2019 pandemic: perspectives from around the world. Gastroenterology 2020 Jul;159(1):26-29 [FREE Full text] [doi: 10.1053/j.gastro.2020.05.012] [Medline: 32387540]
13. Pérez M, Pesek I, Zmazek B, Lipovec A. Video explanations as a useful digital source of education in the Covid 19 situation. J Elem Educ 2020 Nov 12;13(4):395-412. [doi: 10.18690/rei.13.4.395-412.2020]

14. Hilburg R, Patel N, Ambruso S, Biewald M, Farouk S. Medical education during the coronavirus disease-2019 pandemic: learning from a distance. Adv Chronic Kidney Dis 2020 Sep;27(5):412-417 [FREE Full text] [doi: 10.1053/j.ackd.2020.05.017] [Medline: 33308507]
15. Ray EB, Donohew L. Communication and Health: Systems and Applications. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers; 1989:1-224.
16. Lupton D. Toward the development of critical health communication praxis. Health Commun 1994 Jan;6(1):55-67. [doi: 10.1207/s15327027hc0601_4]
17. Dutta-Bergman M. Theory and practice in health communication campaigns: a critical interrogation. Health Commun 2005;18(2):103-122. [doi: 10.1207/s15327027hc1802_1] [Medline: 16083406]
18. Maibach E, Parrott R. Designing Health Messages: Approaches from Communication Theory and Public Health Practice. Thousand Oaks, CA: Sage Publications Inc; 1995:1-320.
19. Berry D. Health Communication: Theory and Practice. Berkshire, United Kingdom: Open University Press; 2006:1-152.
20. Cassata D. Health communication theory and research: a definitional overview. Ann Int Commun Assoc 2017 Nov 02;4(1):583-589. [doi: 10.1080/23808985.1980.11923826]
21. Costello D. Health communication theory and research: an overview. Ann Int Commun Assoc 2017 Nov 02;1(1):557-567. [doi: 10.1080/23808985.1977.11923706]
22. Kar SB, Alcalay R, Alex S. Health Communication: A Multicultural Perspective. Thousand Oaks, CA: Sage Publications Inc; 2001.
23. Kreuter MW, McClure SM. The role of culture in health communication. Annu Rev Public Health 2004 Apr 01;25(1):439-455. [doi: 10.1146/annurev.publhealth.25.101802.123000] [Medline: 15015929]
24. Speaking of Health: Assessing Health Communication Strategies for Diverse Populations. Washington, DC: The National Academies Press; 2002.
25. Parker R, Gazmararian J. Health literacy: essential for health communication. J Health Commun 2003;8 Suppl 1:116-118. [doi: 10.1080/713851963] [Medline: 14692576]
26. Maibach E, Maxfield A, Ladin K, Slater M. Translating health psychology into effective health communication: the American healthstyles audience segmentation project. J Health Psychol 1996 Jul;1(3):261-277. [doi: 10.1177/135910539600100302] [Medline: 22011991]
27. Kreuter M, Wray R. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:227-232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
28. Noar S, Harrington N, Van Stee SK, Aldrich RS. Tailored health communication to change lifestyle behaviors. Am J Lifestyle Med 2010 Nov 19;5(2):112-122. [doi: 10.1177/1559827610387255]
29. Obregón R, Chitnis K, Morry C, Feek W, Bates J, Galway M, et al. Achieving polio eradication: a review of health communication evidence and lessons learned in India and Pakistan. Bull World Health Organ 2009 Aug;87(8):624-630 [FREE Full text] [doi: 10.2471/blt.08.060863] [Medline: 19705014]
30. Cho H, Salmon C. Unintended effects of health communication campaigns. J Commun 2007;57(2):293-317. [doi: 10.1111/j.1460-2466.2007.00344.x]
31. Hinyard L, Kreuter M. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]
32. Snyder LB, Hamilton MA, Mitchell EW, Kiwanuka-Tondo J, Fleming-Milici F, Proctor D. A meta-analysis of the effect of mediated health communication campaigns on behavior change in the United States. J Health Commun 2004 Jan;9 Suppl 1:71-96. [doi: 10.1080/10810730490271548] [Medline: 14960405]
33. Robinson T, Patrick K, Eng T, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. J Am Med Assoc 1998 Oct 14;280(14):1264-1269. [doi: 10.1001/jama.280.14.1264] [Medline: 9786378]
34. Suggs LS. A 10-year retrospective of research in new technologies for health communication. J Health Commun 2006 Jan;11(1):61-74. [doi: 10.1080/10810730500461083] [Medline: 16546919]
35. Neuhauser L, Kreps G. Rethinking communication in the e-Health era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
36. Huo J, Desai R, Hong Y, Turner K, Mainous A, Bian J. Use of social media in health communication: findings from the Health Information National Trends Survey 2013, 2014, and 2017. Cancer Control 2019;26(1):1073274819841442 [FREE Full text] [doi: 10.1177/1073274819841442] [Medline: 30995864]
37. Broniatowski D, Jamison A, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
38. Hilbert M. The end justifies the definition: the manifold outlooks on the digital divide and their practical usefulness for policy-making. Telecommunications Policy 2011 Sep;35(8):715-736. [doi: 10.1016/j.telpol.2011.06.012]
39. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]

40. Eruchalu C, Pichardo M, Bharadwaj M, Rodriguez C, Rodriguez J, Bergmark R, et al. The expanding digital divide: digital health access inequities during the Covid-19 pandemic in New York City. J Urban Health 2021 Apr;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
41. Dadaczynski K, Tolks D. Digital health communication and health literacy in times of COVID-19. Planning and implementation of a special course of study in health promotion and prevention. GMS J Med Educ 2021;38(1):1-8 [FREE Full text] [doi: 10.3205/zma001427] [Medline: 33659636]
42. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. Psychol Behav Sci Int J Juniper 2017 Mar 9;2(2). [doi: 10.19080/PBSIJ.2017.02.555585]
43. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
44. Bandura A. A social cognitive theory of personality. In: Handbook of Personality. New York: Guilford Press; 1999:154-196.
45. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Aug 26;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
46. Rosenstock I, Strecher V, Becker M. Social learning theory and the health belief model. Health Edu Q 2016 Sep 04;15(2):175-183. [doi: 10.1177/109019818801500203]
47. Janz N, Becker M. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
48. Locke EA. Toward a theory of task motivation and incentives. Organ Behav Hum Perform 1968 May;3(2):157-189. [doi: 10.1016/0030-5073(68)90004-4]
49. Tosi HL, Locke EA, Latham GP. A theory of goal setting and task performance. Acad Manag Rev 1991 Apr;16(2):480. [doi: 10.2307/258875]
50. Koch-Weser S, Bradshaw Y, Gualtieri L, Gallagher S. The Internet as a health information source: findings from the 2007 Health Information National Trends Survey and implications for health communication. J Health Commun 2010;15 Suppl 3:279-293. [doi: 10.1080/10810730.2010.522700] [Medline: 21154099]
51. Feng B, MacGeorge E. The influences of message and source factors on advice outcomes. Commun Res 2010 Jun 09;37(4):553-575. [doi: 10.1177/0093650210368258]
52. Aldoory L. Making health communications meaningful for women: factors that influence involvement. J Public Rel Res 2009 Nov 19;13(2):163-185. [doi: 10.1207/S1532754XJPRR1302_3]
53. Balog-Way D, McComas K, Besley J. The evolving field of risk communication. Risk Anal 2020 Nov 20;40(S1):2240-2262 [FREE Full text] [doi: 10.1111/risa.13615] [Medline: 33084114]
54. Berg SH, O'Hara JK, Shortt MT, Thune H, Brønnick KK, Lungu DA, et al. Health authorities' health risk communication with the public during pandemics: a rapid scoping review. BMC Public Health 2021 Jul 15;21(1):1401 [FREE Full text] [doi: 10.1186/s12889-021-11468-3] [Medline: 34266403]
55. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
56. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. Br Med J 2005 Nov 05;331(7524):1064-1065 [FREE Full text] [doi: 10.1136/bmj.38636.593461.68] [Medline: 16230312]
57. Jalali S, Wohlin C. Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. 2012 Presented at: ESEM '12: 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement; September 19 - 20, 2012; Lund Sweden p. 29-38. [doi: 10.1145/2372251.2372257]
58. Peters M, Godfrey C, McInerney P, Munn Z, Trico A, Khalil H. Chapter 11: scoping reviews. In: JBI Manual for Evidence Synthesis. Australia: JBI - The University of Adelaide; 2020.
59. Blelloch G, Hardwick J, Sipelstein J, Zagha M, Chatterjee S. Implementation of a portable nested data-parallel language. J Parall Distrib Comput 1994 Apr;21(1):4-14. [doi: 10.1006/jpdc.1994.1038]
60. Hsieh H, Shannon S. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
61. Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. J Appl Commun Res 2009 May 21;27(3):258-284. [doi: 10.1080/00909889909365539]
62. Hickey G, Kipping C. A multi-stage approach to the coding of data from open-ended questions. Nurs Res 1996 Oct 01;4(1):81-91. [doi: 10.7748/nr.4.1.81.s9] [Medline: 27707369]
63. Finfgeld-Connett D. Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. Qual Res 2013 Apr 04;14(3):341-352. [doi: 10.1177/1468794113481790]
64. Syrjala K, Abrams J, Polissar N, Hansberry J, Robison J, DuPen S, et al. Patient training in cancer pain management using integrated print and video materials: a multisite randomized controlled trial. Pain 2008 Mar;135(1-2):175-186 [FREE Full text] [doi: 10.1016/j.pain.2007.10.026] [Medline: 18093738]
65. Romer D, Jamieson PE, Jamieson KH, Jones C, Sherr S. Counteracting the influence of peer smoking on YouTube. J Health Commun 2017 Apr;22(4):337-345 [FREE Full text] [doi: 10.1080/10810730.2017.1290164] [Medline: 28306481]

66. Curbow B, Fogarty LA, McDonnell KA, Chill J, Scott LB. The role of physician characteristics in clinical trial acceptance: testing pathways of influence. J Health Commun 2006 Mar;11(2):199-218. [doi: 10.1080/10810730500526703] [Medline: 16537288]
67. Visser LN, Tollenaar MS, van Doornen LJ, de Haes HC, Smets EM. Does silence speak louder than words? The impact of oncologists' emotion-oriented communication on analogue patients' information recall and emotional stress. Patient Educ Couns 2019 Jan;102(1):43-52. [doi: 10.1016/j.pec.2018.08.032] [Medline: 30224275]
68. Prieto-Pinto L, Lara-Díaz MF, Garzón-Orjuela N, Herrera D, Páez-Canro C, Reyes JH, et al. Effectiveness assessment of maternal and neonatal health video clips in knowledge transfer using neuromarketing tools: a randomized crossover trial. PLoS One 2019;14(5):e0215561 [FREE Full text] [doi: 10.1371/journal.pone.0215561] [Medline: 31067282]
69. Cox CE, White DB, Abernethy AP. A universal decision support system. Addressing the decision-making needs of patients, families, and clinicians in the setting of critical illness. Am J Respir Crit Care Med 2014 Aug 15;190(4):366-373. [doi: 10.1164/rccm.201404-0728cp]
70. Epstein AS, Shuk E, O'Reilly EM, Gary KA, Volandes AE. 'We have to discuss it': cancer patients' advance care planning impressions following educational information about cardiopulmonary resuscitation. Psychooncology 2015 Dec 23;24(12):1767-1773 [FREE Full text] [doi: 10.1002/pon.3786] [Medline: 25708116]
71. Danila MI, Outman RC, Rahn EJ, Mudano AS, Thomas TF, Redden DT, et al. A multi-modal intervention for Activating Patients at Risk for Osteoporosis (APROPOS): rationale, design, and uptake of online study intervention material. Contemp Clin Trials Commun 2016 Dec 15;4:14-24 [FREE Full text] [doi: 10.1016/j.conctc.2016.06.010] [Medline: 27453960]
72. Withers GF, Twigg K, Wertheim EH, Paxton SJ. A controlled evaluation of an eating disorders primary prevention videotape using the Elaboration Likelihood Model of Persuasion. J Psychosom Res 2002 Nov;53(5):1021-1027. [doi: 10.1016/s0022-3999(02)00493-2]
73. Sacks L, Nakaji M, Harry KM, Oen M, Malcarne VL, Sadler GR. Testicular cancer knowledge among deaf and hearing men. J Cancer Educ 2013 Sep 28;28(3):503-508 [FREE Full text] [doi: 10.1007/s13187-013-0493-x] [Medline: 23813488]
74. McKenzie LB, Roberts KJ, Collins CL, Clark RM, Smith KC, Manganello J. Maternal knowledge, attitudes, and behavioral intention after exposure to injury prevention recommendations in the news media. J Health Commun 2019;24(7-8):625-632. [doi: 10.1080/10810730.2019.1646357] [Medline: 31378155]
75. Bekalu MA, Bigman CA, McCloud RF, Lin LK, Viswanath K. The relative persuasiveness of narrative versus non-narrative health messages in public health emergency communication: evidence from a field experiment. Prev Med 2018 Jun;111:284-290 [FREE Full text] [doi: 10.1016/j.ypmed.2017.11.014] [Medline: 29154794]
76. Phelan EA, Deyo RA, Cherkin DC, Weinstein JN, Ciol MA, Kreuter W, et al. Helping patients decide about back surgery: a randomized trial of an interactive video program. Spine (Phila Pa 1976) 2001 Jan 15;26(2):206-212. [doi: 10.1097/00007632-200101150-00016] [Medline: 11154542]
77. Grindel C, Brown L, Caplan L, Blumenthal D. The effect of breast cancer screening messages on knowledge, attitudes, perceived risk, and mammography screening of African American women in the rural south. Oncol Nurs Forum 2007 Feb 8;31(4):801-808. [doi: 10.1188/04.onf.801-808]
78. Hickey S, Merz EL, Malcarne VL, Gunsauls DC, Huang J, Sadler GR. Breast cancer education for the deaf community in American sign language. Oncol Nurs Forum 2013 Apr 24;40(3):86-91. [doi: 10.1188/13.onf.e86-e91]
79. Berry DL, Halpenny B, Hong F, Wolpin S, Lober WB, Russell KJ, et al. The Personal Patient Profile-Prostate decision support for men with localized prostate cancer: a multi-center randomized trial. Urol Oncol 2013 Oct;31(7):1012-1021 [FREE Full text] [doi: 10.1016/j.urolonc.2011.10.004] [Medline: 22153756]
80. Albert NM, Dinesen B, Spindler H, Southard J, Bena JF, Catz S, et al. Factors associated with telemonitoring use among patients with chronic heart failure. J Telemed Telecare 2016 Jul 08;23(2):283-291. [doi: 10.1177/1357633x16630444]
81. Engler J, Adami S, Adam Y, Keller B, Repke T, Fügemann H, et al. Using others' experiences. Cancer patients' expectations and navigation of a website providing narratives on prostate, breast and colorectal cancer. Patient Educ Couns 2016 Aug;99(8):1325-1332. [doi: 10.1016/j.pec.2016.03.015] [Medline: 27067064]
82. Kempenich JW, Willis RE, Fayyadh MA, Campi HD, Cardenas T, Hopper WA, et al. Video-based patient education improves patient attitudes toward resident participation in outpatient surgical care. J Surg Educ 2018 Nov;75(6):61-67. [doi: 10.1016/j.jsurg.2018.07.024] [Medline: 30217778]
83. Nowak GJ, Evans NJ, Wojdynski BW, Ahn SJG, Len-Rios ME, Carera K, et al. Using immersive virtual reality to improve the beliefs and intentions of influenza vaccine avoidant 18-to-49-year-olds: considerations, effects, and lessons learned. Vaccine 2020 Jan 29;38(5):1225-1233. [doi: 10.1016/j.vaccine.2019.11.009] [Medline: 31806533]
84. Van Alstyne JA, Nowak GJ, Aikin AL. What is 'confidence' and what could affect it?: a qualitative study of mothers who are hesitant about vaccines. Vaccine 2018 Oct 22;36(44):6464-6472. [doi: 10.1016/j.vaccine.2017.09.007] [Medline: 28899629]
85. Otto-Salaj LL, Traxel N, Brondino MJ, Reed B, Gore-Felton C, Kelly JA, et al. Reactions of heterosexual African American men to women's condom negotiation strategies. J Sex Res 2010 Nov 02;47(6):539-551 [FREE Full text] [doi: 10.1080/00224490903216763] [Medline: 19760529]

86. Dudley MZ, Limaye RJ, Omer SB, O'Leary ST, Ellingson MK, Spina CI, et al. Factors associated with referring close contacts to an app with individually-tailored vaccine information. Vaccine 2020 Mar 17;38(13):2827-2832. [doi: 10.1016/j.vaccine.2020.02.019] [Medline: 32098739]
87. Liu Q, Wang Y, Yao MZ, Tang Q, Yang Y. The effects of viewing an uplifting 360-degree video on emotional well-being among elderly adults and college students under immersive virtual reality and smartphone conditions. Cyberpsychol Behav Soc Netw 2020 Mar;23(3):157-164. [doi: 10.1089/cyber.2019.0273] [Medline: 31663773]
88. Hillen MA, de Haes HC, Verdam MG, Smets EM. Does source of patient recruitment affect the impact of communication on trust? Patient Educ Couns 2014 May;95(2):226-230. [doi: 10.1016/j.pec.2014.01.008] [Medline: 24534265]
89. McQueen A, Kreuter MW. Women's cognitive and affective reactions to breast cancer survivor stories: a structural equation analysis. Patient Educ Couns 2010 Dec;81 Suppl:15-21 [FREE Full text] [doi: 10.1016/j.pec.2010.08.015] [Medline: 20850258]
90. Palacio AM, Vidot DC, Tamariz LJ, Uribe C, Hazel-Fernandez L, Li H, et al. Can we identify minority patients at risk of nonadherence to antiplatelet medication at the time of coronary stent placement? J Cardiovasc Nurs 2017;32(6):522-529. [doi: 10.1097/jcn.0000000000000387]
91. Butalid L, Verhaak P, Boeije H, Bensing J. Patients' views on changes in doctor-patient communication between 1982 and 2001: a mixed-methods study. BMC Fam Pract 2012 Aug 08;13:80 [FREE Full text] [doi: 10.1186/1471-2296-13-80] [Medline: 22873783]
92. van Het Schip C, Cheung KL, Vluggen S, Hoving C, Schaper NC, de Vries H. Spoken animated self-management video messages aimed at improving physical activity in people with type 2 diabetes: development and interview study. J Med Internet Res 2020 Apr 23;22(4):e15397 [FREE Full text] [doi: 10.2196/15397] [Medline: 32324138]
93. von dem Knesebeck O, Hoehne A, Link C, Marceau L, Adams A, Roland M, et al. Talking about smoking in primary care medical practice--results of experimental studies from the US, UK and Germany. Patient Educ Couns 2012 Oct;89(1):51-56 [FREE Full text] [doi: 10.1016/j.pec.2012.04.011] [Medline: 22595655]
94. Deyo RA, Cherkin DC, Weinstein J, Howe J, Ciol M, Mulley AG. Involving patients in clinical decisions: impact of an interactive video program on use of back surgery. Med Care 2000 Sep;38(9):959-969. [doi: 10.1097/00005650-200009000-00009] [Medline: 10982117]
95. Geary CW, Burke HM, Johnson L, Liku J, Castelnau L, Neupane S, et al. Personal involvement of young people in HIV prevention campaign messages: the role of message format, culture, and gender. Health Educ Behav 2008 Apr;35(2):190-206. [doi: 10.1177/1090198106288252] [Medline: 17114332]
96. Hannawa AF. "Explicitly implicit": examining the importance of physician nonverbal involvement during error disclosures. Swiss Med Wkly 2012;142:w13576 [FREE Full text] [doi: 10.4414/smw.2012.13576] [Medline: 22573515]
97. Lutfey KE, Gerstenberger E, McKinlay JB. Physician styles of patient management as a potential source of disparities: cluster analysis from a factorial experiment. Health Serv Res 2013 Jun 22;48(3):1116-1134 [FREE Full text] [doi: 10.1111/1475-6773.12005] [Medline: 23088446]
98. Morrongiello BA, Zdzieborski D, Sandomierski M, Lasenby-Lessard J. Video messaging: what works to persuade mothers to supervise young children more closely in order to reduce injury risk? Soc Sci Med 2009 Mar;68(6):1030-1037. [doi: 10.1016/j.socscimed.2008.12.019] [Medline: 19157670]
99. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, et al. The mechanisms and processes of connection: developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. BMC Psychiatry 2019 Dec 21;19(1):413 [FREE Full text] [doi: 10.1186/s12888-019-2405-z] [Medline: 31864315]
100. Shi Z, Wang A, Aronowitz CA, Cappella JN, Romer D, Langleben DD. Connectivity between visual and auditory cortices mediates the influence of argument strength on the effectiveness of smoking-cessation videos among smokers low in sensation seeking. Psychol Res Behav Manag 2019 Jul;12:531-542. [doi: 10.2147/prbm.s183394]
101. Savoia E, Lin L, Viswanath K. Communications in public health emergency preparedness: a systematic review of the literature. Biosecur Bioterror 2013 Sep;11(3):170-184 [FREE Full text] [doi: 10.1089/bsp.2013.0038] [Medline: 24041193]
102. Keller PA, Lehmann DR. Designing effective health communications: a meta-analysis. J Public Policy Mark 2008 Sep 01;27(2):117-130. [doi: 10.1509/jppm.27.2.117]
103. Beech JR, Whittaker J. What is the female image projected by smoking? Int J Psychol Sci 2001;44(3):230-236. [doi: 10.2117/psysoc.2001.230]
104. Meyers-Levy J. The influence of sex roles on judgment. J Consum Res 1988 Mar;14(4):522. [doi: 10.1086/209133]
105. Smith KH, Stutts MA. Effects of short-term cosmetic versus long-term health fear appeals in anti-smoking advertisements on the smoking behaviour of adolescents. J Consum Behav 2006 Dec;3(2):157-177. [doi: 10.1002/cb.130]
106. Fruin DJ, Pratt C, Owen N. Protection motivation theory and adolescents' perceptions of exercise1. J Appl Social Pyschol 1992 Jan;22(1):55-69. [doi: 10.1111/j.1559-1816.1992.tb01521.x]
107. Pechmann C, Shih C. Smoking scenes in movies and antismoking advertisements before movies: effects on youth. J Mark 2018 Nov 28;63(3):1-13. [doi: 10.1177/002224299906300301]
108. Shin S, Chow C, Camacho-Gonsalves T, Levy RJ, Allen IE, Leff HS. A meta-analytic review of racial-ethnic matching for African American and Caucasian American clients and clinicians. J Counsel Psychol 2005;52(1):45-56. [doi: 10.1037/0022-0167.52.1.45]

109. Kim C, McEwen LN, Piette JD, Goewey J, Ferrara A, Walker EA. Risk perception for diabetes among women with histories of gestational diabetes mellitus. Diabetes Care 2007 Sep 15;30(9):2281-2286. [doi: 10.2337/dc07-0618] [Medline: 17575087]
110. Shi Z, Wang A, Aronowitz CA, Romer D, Langleben DD. Individual differences in the processing of smoking-cessation video messages: an imaging genetics study. Biol Psychol 2017 Sep;128:125-131 [FREE Full text] [doi: 10.1016/j.biopsycho.2017.07.019] [Medline: 28757070]
111. Finucane M. Emotion, affect, and risk communication with older adults: challenges and opportunities. J Risk Res 2008;11(8):983-997 [FREE Full text] [doi: 10.1080/13669870802261595] [Medline: 19169420]
112. Damasio A. Descartes' Error: Emotion, Reason, and the Human Brain. London, UK: Penguin Books; 1994:1-336.
113. Lang P. Emotion and motivation: attention, perception, and action. J Sport Exerc Psychol 2016;22(s1):122-140. [doi: 10.1123/jsep.22.s1.s122]
114. Eisenhardt KM. Building theories from case study research. Acad Manag Rev 1989 Oct;14(4):532. [doi: 10.2307/258557]
115. Alvesson M, Kärreman D. Constructing mystery: empirical matters in theory development. Acad Manag Rev 2007 Oct;32(4):1265-1281. [doi: 10.5465/amr.2007.26586822]
116. Sucharew H, Macaluso M. Progress notes: methods for research evidence synthesis: the scoping review approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]

Abbreviations

I-Change: Integrated-Change
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SES: socioeconomic status

Edited by R Kukafka; submitted 04.06.21; peer-reviewed by F Hamilton, P Schulz, CY Lin; comments to author 02.09.21; revised version received 21.09.21; accepted 29.10.21; published 30.12.21.

Please cite as:
Lungu DA, Røislien J, Wiig S, Shortt MT, Ferrè F, Berg SH, Thune H, Brønnick KK
The Role of Recipient Characteristics in Health Video Communication Outcomes: Scoping Review
J Med Internet Res 2021;23(12):e30962
URL: https://www.jmir.org/2021/12/e30962
doi:10.2196/30962
PMID:34967758

Review

Older Adults' Loneliness, Social Isolation, and Physical Information and Communication Technology in the Era of Ambient Assisted Living: A Systematic Literature Review

Rita Latikka[1], MSocSci; Rosana Rubio-Hernández[2], PhD; Elena Simona Lohan[3], PhD; Juho Rantala[1], BBA, MA; Fernando Nieto Fernández[2], PhD; Arto Laitinen[1], PhD; Atte Oksanen[1], DrSocSci

[1]Faculty of Social Sciences, Tampere University, Tampere, Finland

[2]Faculty of Built Environment, Tampere University, Tampere, Finland

[3]Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

Corresponding Author:
Rita Latikka, MSocSci
Faculty of Social Sciences
Tampere University
Kalevantie 4
Tampere, 33100
Finland
Phone: 358 504313178
Email: rita.latikka@tuni.fi

Abstract

Background: Loneliness and social isolation can have severe effects on human health and well-being. Partial solutions to combat these circumstances in demographically aging societies have been sought from the field of information and communication technology (ICT).

Objective: This systematic literature review investigates the research conducted on older adults' loneliness and social isolation, and physical ICTs, namely robots, wearables, and smart homes, in the era of ambient assisted living (AAL). The aim is to gain insight into how technology can help overcome loneliness and social isolation other than by fostering social communication with people and what the main open-ended challenges according to the reviewed studies are.

Methods: The data were collected from 7 bibliographic databases. A preliminary search resulted in 1271 entries that were screened based on predefined inclusion criteria. The characteristics of the selected studies were coded, and the results were summarized to answer our research questions.

Results: The final data set consisted of 23 empirical studies. We found out that ICT solutions such as smart homes can help detect and predict loneliness and social isolation, and technologies such as robotic pets and some other social robots can help alleviate loneliness to some extent. The main open-ended challenges across studies relate to the need for more robust study samples and study designs. Further, the reviewed studies report technology- and topic-specific open-ended challenges.

Conclusions: Technology can help assess older adults' loneliness and social isolation, and alleviate loneliness without direct interaction with other people. The results are highly relevant in the COVID-19 era, where various social restrictions have been introduced all over the world, and the amount of research literature in this regard has increased recently.

(J Med Internet Res 2021;23(12):e28022) doi:10.2196/28022

KEYWORDS

loneliness; social isolation; older adults; physical information and communication technology; systematic literature review

Introduction

Loneliness and social isolation can occur at any stage of human life. Particular attention has been paid to these circumstances among older adults, an increasing demographic in many societies. In 2050, over 20% of the population of most countries will be over 60 years old [1], and there is a strengthening global trend of those living alone later in their life [2]. During old age, the size of one's social network and the extent of social activities are likely to reduce [3-5]. This topic has also become timely

due to the COVID-19 crisis where various country-level restrictions and governmental recommendations on social distancing have been introduced [6,7]. These factors together bring forth the importance of addressing loneliness and social isolation among older adults.

Loneliness refers to perceived social isolation or a subjective unpleasant and distressing feeling that results from a significant discrepancy or mismatch between one's actual and desired social relationships [8-10]. In the typology of social and emotional loneliness, social loneliness is characterized by the lack of engaging social networks, and emotional loneliness refers to the lack of close emotional attachment [11,12]. Social isolation typically concerns an objectively limited or a lack of social contact with others [8], and some of its common quantifiable markers are a shortage in one's social contacts and network size [13]. Despite their similarities, loneliness and social isolation are not the same [12,14]. Loneliness is a subjective emotional feeling, whereas social isolation describes an objective and a quantifiable aspect of social relationships [13]. For instance, the quality of social relationships is more closely related to loneliness compared to the quantity of social relationships [12,15,16].

Recent prevalence estimates from the United States show that more than 40% of older adults are lonely (29% occasionally and 19% frequently) [8,17]. In Europe, prevalence estimates range from Central and Eastern Europe's 30%-55% to Northwestern Europe's 10%-20% [18,19]. As for social isolation, a recent estimate considers 24% of older adults aged 65 and above as socially isolated [20]. Various demographics can also be used for prevalence estimation purposes [21]. Prevalence estimates fluctuate across research studies due to differences in the considered populations, measures, age groups, sample sizes [22], definitions, intensity and duration of the experience [23], and cultural differences [24]. The extent to which people are willing to self-report their loneliness experiences needs to be critically considered. For example, research indicates that men are more reluctant to admit their loneliness than women, likely due to a stigma associated with it [25]. However, instead of asking about loneliness directly, indirect validated measures can also be applied [17].

Loneliness and social isolation are significant predictors of mortality [13,14,26,27], and they are associated with poorer physical and mental health [8]. For instance, loneliness is associated with poorer cardiovascular health [28], lower cognitive function [29,30], depression [31,32], anxiety, suicide ideation [33], higher psychological distress [34,35], lower self-esteem [36], sleep and stress problems [37,38], and health behaviors such as lower physical activity [39]. In turn, social isolation is associated with lower self-rated physical health [40], lower health-related quality of life and health status [41], worse cardiovascular and mental health [42], and increased vulnerability to dementia [43]. Therefore, it is evident that solutions to combat both circumstances are needed.

Partial solutions to assess loneliness and social isolation among older adults have been sought from the field of information and communication technology (ICT). Previous literature reviews have examined empirical studies on various types of technologies and their effectiveness in alleviating social isolation [44]. Other reviews address interventions targeting loneliness and social isolation among older people, which include technological and nontechnological approaches [45,46]. There are also reviews on diverse technologies and caregiving that have identified their impact on loneliness and social isolation alleviation, among other effects [47,48]. These studies address loneliness and social isolation from the perspective of fostering social networking and support, together with community interaction and engagement.

However, we assume that there are habits other than communication with other humans that can also be related to loneliness and social isolation, and these habits can be assessed and tracked using novel intelligent technologies. In particular, robots, wearables, and smart homes hold potential value in this area. In this review, these technologies are grouped under the term "physical ICT," broadly referring to physical technologies able to collect and communicate information. Robots are viewed as embedded agents that can interact with humans or with other robots in a socially acceptable manner, also known as social robots [49,50]. Wearables refer to technologies that can be worn on the human body, such as virtual reality (VR) headsets, fitness trackers, smart watches, or smart jewelry. The term "smart home" (or "smart house") refers to a residence equipped with "smart technology," namely a variety of internet-connected sensors and systems enabling monitoring and management to automate and optimize control of the home environment, home appliances, and the inhabitant's quality of life [51].

Previous reviews have also generally reflected the effectiveness of social robots in elderly care, including studies that address loneliness or social isolation [52,53], and the influence of smart houses on older adults' quality of life, including their effect on social isolation [54]. Recently, in the face of the COVID-19 pandemic, there has also been a stream of studies from different fields reviewing and considering the importance and possibilities of robots and computer agents in alleviating loneliness [55-58]. To the best of our knowledge, no prior studies have focused on role of physical ICT solutions in assessing and combating loneliness and social isolation among older adults.

When the solutions relate to health and care, the concept of ambient assisted living (AAL) comes into play. AAL is a subarea of ambient intelligence and can be defined as "an emerging multidisciplinary field aimed at providing an ecosystem of different types of sensors, computers, mobile devices, wireless networks, and software applications for personal health care monitoring and telehealth systems" [59]. AAL was first coined in 2006 by the International Medical Informatics Association in recognition of this emerging technology with the creation of a working group on smart homes and AAL [60].

The aim of this study is to gain insight into how physical ICTs can help overcome loneliness and social isolation among older adults other than by fostering social communication with people and what the main open-ended challenges according to the reviewed studies are. Our focus is on empirical research conducted from January 2006 to late May 2021, starting from

the year in which the concept of AAL was introduced. In line with these aims, we established the following research questions:

> *(RQ1) What has been studied so far, from a sociotechnological perspective, in the field of loneliness and social isolation in older adults using physical ICT solutions?*
>
> *(RQ2) How can physical ICT solutions help overcome the issues of loneliness and social isolation among older adults other than by fostering social communication with people?*
>
> *(RQ3) What are the main open-ended challenges according to existing studies?*

Methods

Data Collection

A systematic literature review was conducted to answer our research questions. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedure [61] was followed, when applicable, for the study objectives. The data were collected in 2 phases. The first phase took place in April 2020, covering the period from January 2006 to late March 2020. The second phase was conducted in June 2021, covering the time frame from April 2020 to the end of May 2021, which allowed us to keep the data up to date.

In both phases, the procedure was the same. We used 7 bibliographic databases: Scopus (Elsevier), Web of Science (Clarivate), EBSCOhost (EBSCO), Social Science Premium Collection (ProQuest), PsycINFO (Ovid), PubMed (National Library of Medicine), and IEEE Xplore Digital Library (IEEE), with all databases selected. The following search phrases were used in the databases: (“ambient assisted living” OR “ambient intelligence” OR “smart house” OR “smart home” OR “smart environment” OR “smart assistant*” OR “intelligent assistant*” OR sensor* OR “internet of things” OR wearable* OR robot* OR “artificial intelligence”) AND (eld* OR age* OR old* OR geriatr* OR senior*) AND (lone* OR “social* isolat*”).

The search was targeted toward the “title,” “abstract,” and “keywords.” In PsychINFO, “key concepts” were selected as corresponding to keywords. In Social Science Premium Collection, “all subjects and indexing” including keywords and index terms was the term used in addition to abstracts and document titles. EBSCOhost and Social Science Premium Collection searches were filtered to include only peer-reviewed entries, and IEEE Xplore Digital Library was filtered for conference and journal publications to manage the number of irrelevant entries. All searches were limited to English language publications.

In the first phase, the search from the 7 different databases first produced 1830 results in total. After removing duplicate results, the data consisted of 1001 entries. In the second phase, the search resulted in 559 entries. After removing the duplicates and overlaps with data from the first phase, the additional data consisted of 270 entries. In both phases, all papers were screened according to the predefined inclusion criteria. The studies were included based on the following criteria:

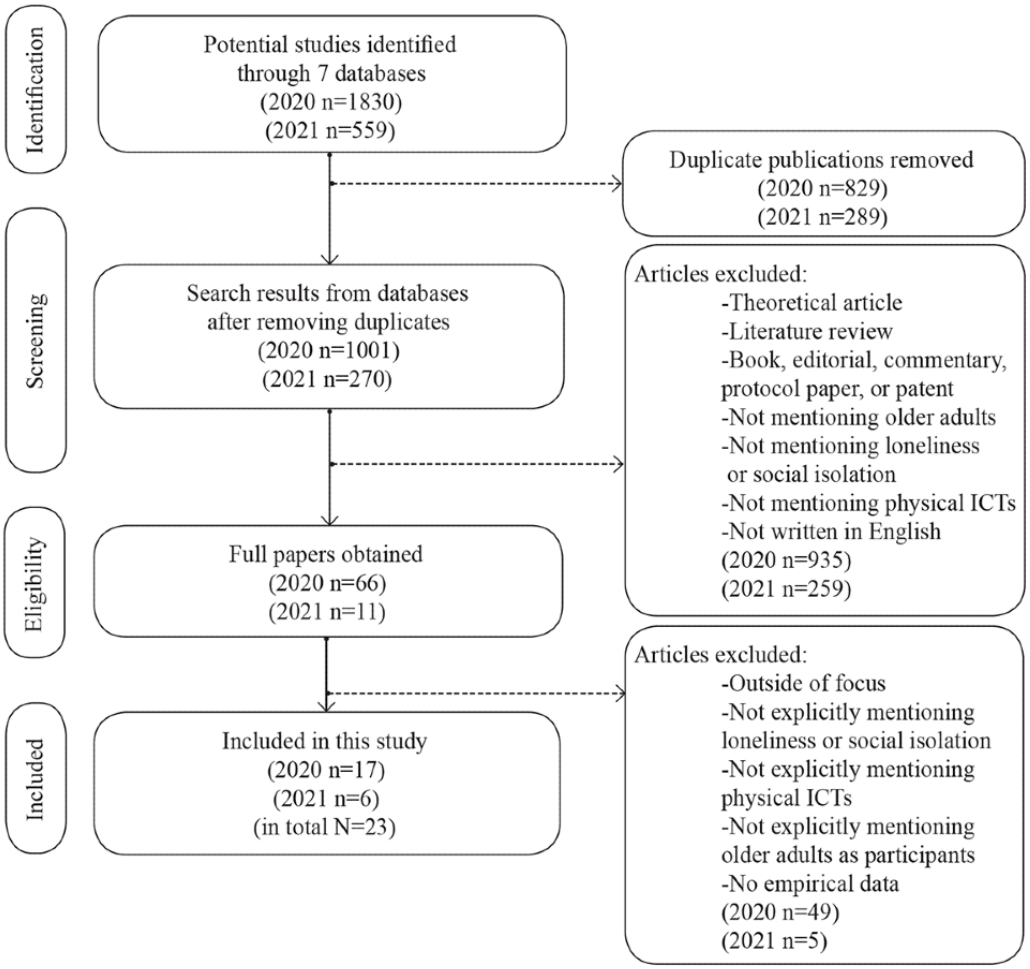
> *(C1) It is a peer-reviewed article or a conference publication.*
>
> *(C2) It is a fully empirical study, using quantitative, qualitative, or mixed methods.*
>
> *(C3) The study does fully or partially research older adults.*
>
> *(C4) The study does fully or partially research human loneliness or social isolation.*
>
> *(C5) The study does fully or partially research physical ICTs (robots, wearables, or smart homes and houses).*

Consequently, we focused on empirical studies in which physical ICT solutions are researched with older adults and loneliness or social isolation is explicitly examined. To be included, ICT solutions had to be physically exploited and studied in relation to older adults’ loneliness and social isolation, and not solely aimed at mediating or fostering communication between people. All study participants referred to as older adults in the selected studies were considered eligible for our study purposes (starting from people over 50 years old). Due to a technology-focused research topic, all relevant studies were searched, including peer-reviewed journal and book articles, as well as conference publications. Studies were excluded if they were theoretical articles or literature reviews; if they were whole books, editorials, commentaries, study protocols, or patents; if they did not explicitly mention older adults, loneliness, social isolation, or any physical ICT solution; or if the full text was not written in English.

In the first search phase, 2 coders first independently reviewed the papers’ titles, abstracts, and keywords, after which selections based on the predefined inclusion criteria were made. An interrater reliability test was conducted and resulted in an interrater agreement of 94.57% on average (mean Cohen κ=0.83, range 0.74-0.88), indicating a successful set of criteria. Borderline cases and disagreements were discussed with the members of the research team. Then, 66 full papers were screened by 2 coders against the predefined inclusion criteria, of which 17 papers were included in the data set.

In the second search phase, we extended data collection until the end of May 2021, and the same procedure was followed. An interrater reliability test resulted in an interrater agreement of 94.67% on average (mean Cohen κ=0.79, range 0.48-1.00), replicating the success of the set criteria. Further, 11 full papers were independently screened by 2 coders, of which 6 papers were included in the data set. Hence, the final data set consisted of 23 empirical studies. The diagram depicting the entire data collection and data selection process is presented in Figure 1.

Figure 1. Diagram of the entire data collection and data selection process including both search phases. ICTs: information and communication technologies.



Method of Analysis

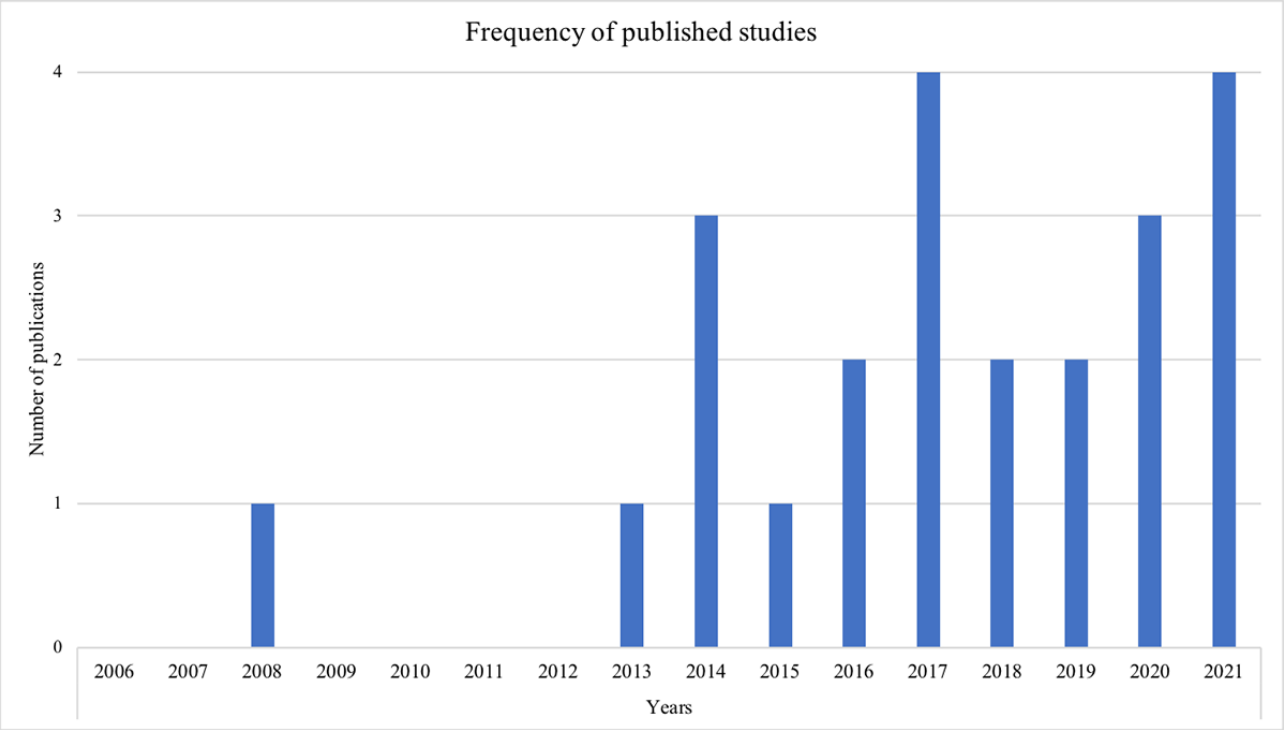
We began the analysis by examining what has been studied from a sociotechnological perspective in the field of loneliness and social isolation in older adults using physical ICT solutions. A descriptive overview of the studies was obtained to include the following basic characteristics of the studies coded into an Excel sheet (Microsoft Corporation): research method (quantitative, qualitative, or mixed), type of study, time frame, study setting, older adults' sample size, age, and gender information, research instrument for measuring loneliness or social isolation, type of physical ICT addressed, and focus of the paper (detection or prediction, alleviation, or other). Then, content analysis was performed to summarize the ways in which new technologies can help overcome issues of loneliness and social isolation among older adults and the main open-ended challenges according to the included studies. Meta-analysis was not conducted due to the heterogeneity of the reviewed technologies and methods used in these studies. Furthermore, our methodological choices were limited by the small number of existing studies on the topic.

Results

Descriptive Details of the Reviewed Studies

The final data set consisted of 23 studies conducted from 2006 to the end of May 2021. Except for 1 study, all the others were published after 2012. This suggests that the amount of research has increased over the past decade, as depicted in Figure 2. Studies have mostly been conducted in the United States (n=10), Germany (n=2), and Singapore (n=2). Other countries include Australia, Canada, Ireland, Mexico, the Netherlands, New Zealand, and Taiwan. Among these, 2 studies included cross-national data, 1 with participants from the United Kingdom, Italy, and Ireland, and another with participants from England and Japan. Most of the studies have applied quantitative (n=13) or mixed methods (n=7), whereas very few studies have applied qualitative methods (n=3). Most of the included studies contributed to the field of older adults' loneliness (n=17), whereas the rest examined social isolation or both phenomena (n=6).

Figure 2. Frequency of publications per year from 2006 to the end of May 2021.



Main Areas of Research

The 2 main areas of the reviewed research comprised "detection and prediction" and "alleviation" of older adults' loneliness and social isolation using physical ICT. Here, the category of detection and prediction refers to studies that hypothesize the potential of certain activities or daily habits (attributes), including time spent in certain rooms of the house and outdoors, to infer the levels of loneliness or social isolation in older adults. In the reviewed studies, certain daily activities are tracked in real time through physical ICTs, namely smart home solutions, or a combination of smart homes and smartphones. Then, algorithms derive behavioral patterns from the gathered data. Afterward, these factual scores are correlated with subjective standard measurements to identify the attributes that meaningfully relate to loneliness or social isolation. In addition, studies validate their predictive models using different evaluation methods and indexes.

Alleviation studies include examining whether the use of technology would result in reducing perceived loneliness or social isolation in older adults, or the possible roles physical ICT could play in combating older adults' loneliness and social isolation, if not serving as an actual intervention. Further, 1 paper was categorized as "other," referring to a study examining the association between perceived loneliness and acceptance of robots. It was also included in the final data set, as it has implications in terms of overcoming loneliness from an opposite perspective; without acceptance of the intervention for reducing perceived loneliness, there is no possible cure.

The focus was on detection and prediction in less than half of the studies (n=7), whereas the alleviation of loneliness or social isolation was the focus in most of the considered studies (n=15). None of the studies directly examined both aspects. The basic information of the included studies is presented in Table 1.

Table 1. Basic information of the selected studies (N=23).

Method and context	Focus			
	Detection, n (%)	Alleviation, n (%)	Other, n (%)	Total, N (%)
Method				
Quantitative	6 (85.71)	6 (40)	1 (100)	13 (56.52)
Qualitative	—[a]	3 (20)	—	3 (13.04)
Mixed method	1 (14.29)	6 (40)	—	7 (30.44)
Total	7 (100)	15 (100)	1 (100)	23 (100)
Study context and countries				
United States	3 (42.86)	7 (46.67)	—	10 (43.48)
Singapore	2 (28.57)	—	—	2 (8.7)
Germany	—	1 (6.67)	1 (100)	2 (8.7)
Australia	—	1 (6.67)	—	1 (4.35)
Canada	—	1 (6.67)	—	1 (4.35)
Ireland	1 (14.29)	—	—	1 (4.35)
Mexico	1 (14.29)	—	—	1 (4.35)
The Netherlands	—	1 (6.67)	—	1 (4.35)
New Zealand	—	1 (6.67)	—	1 (4.35)
Taiwan	—	1 (6.67)	—	1 (4.35)
United Kingdom, Italy, and Ireland	—	1 (6.67)	—	1 (4.35)
England and Japan	—	1 (6.67)	—	1 (4.35)
Total	7 (100)	15 (100)	1 (100)	23 (100)

[a]—:not available

Types of Technologies Assessed

Within the reviewed studies on the detection and prediction of loneliness and social isolation, platforms with several types of sensor-based technologies were commonly used, namely, smart home solutions. Austin et al [62], Goonawardene et al [63], Petersen et al [64,65], and Walsh et al [66] assessed smart home solutions explicitly, whereas Huynh et al [67] analyzed sensor-enabled homes. Martinez et al [68] combined smart homes and smartphones in their study. In the reviewed studies focusing on loneliness and social isolation alleviation, older adults have been introduced most often to social robots [69-79]. Studies have also examined the use of a smart home solution [80], VR systems [81,82], and an ambient activity system that includes an activity sensor (AAL-VU system) [83]. Furthermore, 1 study focusing on how loneliness associates with physical ICT acceptance exploited social robots Paro and Giraff [84].

Study Designs and Settings

The reviewed studies applied various study designs ranging from explorative pilot studies to randomized controlled experiments. Longitudinal design was applied in some studies [62,65], and 3 studies followed the randomized controlled trial protocol [69,76,77]. A majority of the 23 studies applying quantitative measurements used validated scales for subjectively measuring loneliness and social isolation: the original, revised, or short version of the University of California Los Angeles (UCLA) Loneliness Scale (n=12), Dong Jong Gierveld Loneliness Scale (n=4), or Lubben Social Network Scale (n=2). However, most studies were conducted with relatively small sample sizes, which is also understandable due to the use of new technologies and the experimental nature of the studies. Many studies targeted healthy older adults with no cognitive impairment with the exceptions of Appel et al [81], focusing on older adults with cognitive and physical impairments; Robinson et al [77], in which 19 out of the 40 participants had cognitive impairment; Fields et al [72], where roughly half of the participants had dementia; and Casey et al [70] focusing on people with dementia. Further, Chen et al [71] focused on older adults with depression, and Hudson et al [74] and Tkatch et al [79] noted that their participants had higher levels of depression compared to the ones who declined from participating in the study. Studies were conducted in the older adults' own homes, different forms of care facilities, sensor-enabled houses and apartments, and hospitals. Table 2 gives a descriptive overview of the considered studies.

Table 2. Descriptive overview of the selected studies (N=23).

Studies and method used	Type of study	Time	Setting	Environment	No. of participants	Age in years	Female (%)	Instrument	Type of technology and focus
Appel et al [81]; mixed	Feasibility	3×20 min[a]	Facility	Indoors	66	Mean 80.5	60. 6	1 item in STAI[b]	VR[c] system; alleviation
Austin et al [62]; quantitative	Longitudinal	8 mo[d]	Home	Indoors	16	>62 (mean 71)	81	UCLA[e]	Smart home; detection
Baisch et al [84]; quantitative	User/field study	N/A[f]	N/A	Indoors	29	65-81 (median 70)	79	1 item	Paro and Giraff robots; other
Banks et al [69]; quantitative	RCT[g]	8 wk[h]	Facility	Indoors	38	N/A	N/A	UCLA	AIBO robot; alleviation
Brandenburgh et al [83]; mixed	Feasibility	6 wk	Home	Indoors	Multip.[i]	Multip.	Multip.	DJGLS[j]	AAL-VU[k] system; alleviation
Casey et al [70]; qualitative	Interview	2 mo	Home, facility, hospital	Indoors	38	Multip.	Multip.	Interview	MARIO robot; alleviation
Chen et al [71]; mixed	Quasi-experiment and interview	8 wk	Facility	Indoors	20	65-93 (mean 81.1)	65	UCLA	Paro robot; alleviation
Curumsing et al [80]; mixed	Case study, field trial	13 wk	Home	Indoors	10	>65	N/A	Interview	Smart home; alleviation
Fields et al [72]; quantitative	Pilot experiment	3×10 min	Facility	Indoors	15	77-92 (mean 85.80)	73.3	UCLA	NAO robot; alleviation
Follman et al [73]; mixed	Experiment and interview	2 mo	Facility, hospital	Indoors	70	59-98 (av.[l] 83)	72.86	UCLA	temi robot; alleviation
Goonawardene et al [63]; mixed	Multimethod	7 mo	Smart home	Indoor/outdoor	46	60-91	58.7	LSNS[m], DJGLS, social activity attendance	Smart home; detection
Hudson et al [74]; qualitative	Interview	1 h[n]	Home	Indoors	20	65-90 (av. 76)	50	Interview	Pet robots; alleviation
Huynh et al [67]; quantitative	Field study	6 mo	Smart home	Indoor/outdoor	43	Mean 77.59	N/A	DJGLS	Smart home; detection
Lazar et al [75]; qualitative	Focus group	N/A	Home, facility	Indoors	41	61-92 (mean 77)	85.37	Interview	Pet robots; alleviation
Lin et al [82]; quantitative	Field study	2 wk	Facility	Indoors	63	Born 1918-1950	62	UCLA	VR system; alleviation
Martinez et al [68]; quantitative	Multimethod	N/A	Home	Indoors/outdoors	Multip.	Multip.	Multip.	LSNS	Smart home; smartphone; detection
Papadopoulos et al [76]; quantitative	RCT	18 h across 2 wk	Facility	Indoors	33	65-98 (mean 81.9)	66.7	UCLA (ULS-8[o])	Pepper robot; alleviation

Studies and method used	Type of study	Time	Setting	Environment	No. of participants	Age in years	Female (%)	Instrument	Type of technology and focus
Petersen et al [64]; quantitative	Multimethod	5 d[p]	Facility, home	Outdoors	Multip.	Multip.	Multip.	UCLA	Smart home; detection
Petersen et al [65]; quantitative	Longitudinal	12 mo	Facility, home	Outdoors	85	65-96 (mean 86.36)	85	UCLA	Smart home; detection
Robinson et al [77]; quantitative	RCT	12 wk	Facility	Indoors	40	55-100	N/A	UCLA	Paro robot; alleviation
Sidner et al [78]; mixed	Field study	1 mo	Home	Indoors	44	55-91 (mean 66)	N/A	UCLA	AlwaysOn, robot/virtual; alleviation
Tkatch et al [79]; quantitative	Intervention	1 mo	Home	Indoors	216	65-85	Multip.	UCLA	Animatronic/robotic pets; alleviation
Walsh et al [66]; quantitative	Field study	4×28 d	Smart home	Indoors	13	60-88	46.15	DJGLS	Smart home; detection

[a]min.: minutes.
[b]STAI: State-Trait Anxiety Inventory.
[c]VR: virtual reality.
[d]mo: month or months.
[e]UCLA: University of California Los Angeles Loneliness Scale.
[f]N/A: not applicable/not mentioned.
[g]RCT: randomized controlled trial.
[h]wk: week or weeks.
[i]Multip.: multiple data.
[j]DJGLS: Dong Jong Gierveld Loneliness Scale.
[k]AAL-VU: an ambient system.
[l]av.: average.
[m]LSNS: Lubben Social Network Scale.
[n]h: hour or hours.
[o]ULS-8: UCLA Loneliness Scale-8.
[p]d: day or days.

Detecting and Predicting Older Adults' Loneliness and Social Isolation via Behavioral Attributes

The studies aiming to overcome loneliness and social isolation through detection and prediction are framed within the assumption that an early diagnosis of loneliness and social isolation in older adults can prevent their physical and psychosociological decline. All 7 papers report relevant results and provide a meaningful quantitative or qualitative evaluation of their experiments. All the solutions are shown, at least to some extent, to be capable of detecting and predicting older adults' loneliness or social isolation; thus, they help in the process of overcoming such circumstances by recognizing the existence of these phenomena. Overall, researchers refer to these detection systems as promising research paths toward overcoming loneliness and social isolation.

In practice, studies assess various older adults' behavioral attributes and compare them with subjective measures. Older adults' out-of-home habits measured in 5 of the studies [63-65,67,68] are reported as relevant attributes in inferring loneliness and social isolation. Thus, tracking attributes related to outings (time spent outside the house, number of outings, and number of places visited) seems consistently relevant in unobtrusive models to detect loneliness or social isolation. Nevertheless, other attributes analyzed in these studies are not as consistent across the studies in terms of their relevance to detection. Different studies report that diverse variables should be considered in prediction models with this type of technology. Time spent in the house is generally relevant in 1 study [68], whereas it is not in another [62]. Other reported significant variables include time spent in the living room [63,66], time spent across various locations [66], walking speed [62], nocturnal movements [66], and daytime napping [63].

Studies vary in terms of the development phase of the system. Martinez et al [68] propose a smartphone app that enables the users and caregivers to interact and receive feedback from the

system, whereas Huynh et al [67] and Walsh et al [66] elaborate on potential interfaces. Some studies expand on the core issue of loneliness and social isolation inference. Huynh et al [67] also assess depression among independently living seniors in their study. Petersen et al [65] focused on monitoring the broad cognitive, physical, and emotional states of older persons, and Walsh et al [66] also explored anxiety, cognition, depression, independent living skills, sleep quality, and quality of life.

Papers also report some study-related issues, such as small sample sizes [62,66,68], study durations [65,67], and nonoptimal study designs [62,65]. Likewise, some studies report technical problems, such as ensuring that sensor configurations work continuously [66], assessing situations with more than one person living in a house [66,67], and the intrusiveness of smartphone models [68]. Moreover, 2 studies note general ethical challenges regarding issues such as privacy, respect, or consent [62,66].

Alleviating Older Adults' Loneliness and Social Isolation Through Physical ICT

Studies on alleviation indicate that some physical ICT solutions can help overcome loneliness among older adults by decreasing it based on self-reported measures. None of the studies report complete elimination of older adults' loneliness experiences with the help of physical ICT, but they report success in alleviating them. Studies that report statistically significant results for a decrease in perceived loneliness of older adults include interventions using social robots [69,71-73,77,79] and an ambient activity system including an activity sensor [83]. Moreover, 2 studies report qualitative evidence of using social robots in alleviating loneliness [70,74]; in addition, 1 study reports unexpected qualitative evidence regarding the unexpected positive impacts of an intervention using a smart home solution on loneliness [80]. No statistically significant results were demonstrated for a decrease in social isolation, but 2 studies suggest that social robots could help combat social isolation by providing social contact with the robots and through video calls with people [70,73].

Due to the variety of physical ICTs employed in the studies on loneliness alleviation, common characteristics regarding intervention designs or technological features are limited. However, a common feature of the successful solutions is that they can interact with the user in one way or another. The successfully employed social robots (Paro, AIBO, NAO, MARIO, temi, and animatronic or robotic pets) are all designed to engage with the user, thus providing means for interaction with the device itself. MARIO and temi robots also provide means for video communication with other people as a central feature besides speech, activity, and entertainment capabilities. In an ambient activity stimulating system [83], a virtual coach gives recommendations to the user based on the planned activities for the day and the data from an activity sensor. The smart home solution [80] also interacts with the user via messages.

Banks et al [69] assessed the mechanism leading to changes in loneliness and report that the intervention did not succeed in alleviating loneliness through attachment to the robotic dog (AIBO) or a living dog. Hudson et al [74] report that many of their participants stated that the presence of the robotic pet positively influenced their feelings of loneliness. Curumsing et al [80] also report that older adults missed the voice of that "somebody" in their home when they uninstalled the smart house system. Casey et al [70] describe that their participants thought that MARIO reduced their feelings of loneliness because the robot provided distractions, allowed them to engage in various activities and interact with family members.

However, the evidence found by these studies must be interpreted cautiously. Studies acknowledge that more valid results would have been obtained with larger sample sizes [70,72,73,76,77], and the lack of control or comparison groups limits the interpretations of the effectiveness of the intervention [71,72,77,79]. Scholars also recognize the possible influence of the presence of a researcher on their results [70,72,76,77] and raise concerns on study dropouts due to older adults' health issues [77]. Robinson et al [77] performed the only study conducting a group session, and they reflect on the possible effects of that group setting on the intervention in a subsequent study [85]. Banks et al [69] noticed that the robotic dog AIBO was not used at full capacity in the study, which may have influenced the results. Further, 6 studies using social robots included participants with cognitive impairment [77], dementia [70,72], and depression [71,74,79], which must also be acknowledged when interpreting the results. Curumsing et al [80] did not focus on alleviating loneliness in their initial study plan, but the topic emerged during the investigation, leading to unexpected results.

Perceived Loneliness in Technology Acceptance

One study shed light on the link between perceived loneliness and acceptance of robots. Baisch et al [84] investigated emotional loneliness as a component of psychosocial functioning and its link with the intention to use social robots Paro and Giraff in 2 different user-technology fit scenarios. According to the results, participants with lower psychological resources, including perceived emotional loneliness, accepted the Giraff robot less when the user-technology fit was poor. The same results were not obtained for a companion robot. However, researchers warn against generalizing the results due to a small and nonrepresentative sample.

Main Open-Ended Challenges According to the Reviewed Studies

The main common open-ended challenges according to the reviewed studies relate to the need for more robust study samples and study designs. Across studies, researchers suggest future studies using larger study samples that could also be more diverse [62,65] and more geographically widespread [77,82], include more people with different conditions [77], and vary more in terms of the living environments [77,82] to add statistical power for generalization, to validate the interpretation of results, and to understand for whom the technologies are best suited. However, researchers admit that realistically obtaining larger samples includes challenges such as the health problems of the seniors [77] and finding isolated adults for recruitment in general [66,78]. Further, it is critical for future studies to incorporate control and comparison groups and confounding factors to the study designs to aid the interpretation of the

effectiveness of the interventions [71,72,76,77,79]. Finally, scholars note that research would benefit from longer study periods [63,65,67,70,73,74,76,84].

Among the detection and prediction studies, Goonawardene et al [63] and Petersen et al [64] propose introducing potentially confounding variables (eg, depression or mobility) and establishing causal relationships between them, applying multivariate estimation models to improve accuracy. For accuracy purposes, Walsh et al [66] recommend tracking the exact day when data are gathered instead of relying on average measurements, and Petersen et al [65] encourage implementing detailed environmental variables (eg, weather conditions, seasonal changes, proximity of resources, ease of transportation, living alone or within a community, or neighborhood demographics).

Regarding the alleviation interventions, more knowledge is needed regarding the effect of group sessions on the intervention results [77,85] to ascertain whether to conduct studies individually or in group settings [72]. The need for advanced and low-cost robotic devices to be used for research purposes [72,75] sets an open challenge for the growing robot market. The study by Lazar et al [75] challenges the perception of seeing pet robots as "technological fixes" for loneliness or social isolation alleviation; instead, it suggests reimagining their potential to suit specific needs and existing social lives. Brandenburgh et al [83] state that future developments may focus on how to change human behavior in addition to which behavior should be changed.

Regarding technological features, Appel et al [81] mention open challenges for future VR studies to succeed better in alleviating feelings of being lonely by applying joint or multiuser experiences with multidisplay setups. Lin et al [82] state that other VR-related systems and technologies could also be examined, but they do not specify the types of systems. Casey et al [70] suggest that robots used in dementia care could have more human-like features and better capabilities for communication and understanding speech in future. Chen et al [71] advocate a reliable method for measuring interaction time, such as an in-built function in Paro. Tkatch et al [79] propose effect comparisons between having a real pet or a robotic pet, and between a robotic cat and a robotic dog. Hudson et al [74] suggest implementing robotic pets with increased interactivity in future interventions; however, they also raise a potential ethical issue where older adults become dependent on their pet. Sidner et al [78] add that instead of using predetermined methods for introducing activities in the system, more flexible and teachable methods could be sought to prevent failures during activity usage in future.

Baisch et al [84] recommend more comprehensive research on the psychological mechanisms and human characteristics with respect to the acceptance of robots, as well as the diversity of older adults' life circumstances. The stage is also open for new ideas to enhance the perceptions of user-technology fit among older adults.

Discussion

Principal Results

This systematic literature review examined the research conducted on older adults' loneliness and social isolation and physical ICT solutions in the era of AAL. The aim was to gain insight into how technologies can help overcome such circumstances without fostering social communication with people and what the main open-ended challenges according to the included studies are. The results demonstrate that issues of loneliness and social isolation among older adults cannot be eliminated using physical ICTs, but that physical ICTs are used to help detect and predict, or alleviate such circumstances. ICT solutions such as smart homes can help predict and detect loneliness and social isolation, and technologies such as robotic pets and some other social robots can help alleviate loneliness to some extent. The main common open-ended challenges according to the reviewed studies are related to the need for more robust study samples and study designs. In addition, studies reported some technology- and topic-specific open-ended challenges.

Based on our findings, all the reviewed smart home and house solutions in the area of detection and prediction were considered capable of detecting and predicting older adults' loneliness or social isolation to some extent. Tracking the outings, including the time spent outside the house, number of outings, and number of places visited, stood out as a relevant activity to be examined in the unobtrusive models. Overall, these methods show a promising research path for overcoming loneliness and social isolation.

Open-ended challenges in the area of detection and prediction included a more flexible adaptation of predictive models to contingencies and contextual situations and the development of learning algorithms that allow the systems to accurately respond to the evolving circumstances of older adults. In addition, ethical issues posed by the intrusiveness of monitoring systems in older adults' lives, as well as economic concerns, remain to be assessed. Further, as pointed out in one of the studies [65], it is relevant that researchers continue to find a way to implement the complexity and variety of living spaces in their experiments if they seek accuracy in their results and solutions to overcome loneliness and social isolation. It is clear that the environment shapes individuals, and thus, these technologies could be exploited further to "attune" [86] the built environment with "typical human situations" [87] to contribute toward increasing the social inclusion of older adults and alleviate feelings of loneliness.

According to the results, some forms of physical ICT hold promising value in alleviating older adults' loneliness. Significant positive results were obtained in interventions using social robots, an ambient activity system that includes an activity sensor, and a smart home solution. A common feature of these solutions is that they are able to interact with the user in one way or another, which may have served as the key to their success by fostering reciprocity between the device and the user, thus providing increased social contact opportunities for older adults. Some studies suggest that social robots could also

help reduce social isolation by providing social contact and activities for older adults, but more robust evidence is still needed to prove that social contact is being increased and not being replaced by technology. One reviewed study demonstrated that the experience of loneliness itself as part of psychological functioning was associated with lower acceptance of a telepresence robot in a poor user-technology fit scenario. Future studies should continue investigating the role of subjective feelings of loneliness in technology acceptance and adoption longitudinally to draw comprehensive conclusions.

Open-ended challenges related to alleviation included uncertainty about whether the interventions should be conducted individually or in group settings and the high cost of technologies. Regarding technological features, such as VR, joint or multiuser experiences with multidisplay setups could be explored in future. For social robots, scholars seek enhanced interaction and speech recognition capabilities, possibly in-built means for measuring interaction time, and further comparisons on the effectiveness of different robotic pets. There is also a need for more flexible and teachable methods for systems aiming to alleviate loneliness and developing knowledge on how technologies can be used for changing human behavior.

Furthermore, we found that none of the solutions in the reviewed studies aimed to alleviate and detect as well as predict older adults' social isolation or loneliness explicitly. Some of the solutions are perhaps able to do so, but none of the studies have examined both aspects by empirical methods. Solutions to assess and combat social isolation have so far been less researched compared to loneliness and are particularly needed. The need is evident in the COVID-19 era with on-going lockdowns and social restrictions all over the world, and where fast digitalization and ICT adoption on large scales have also happened at an unprecedented pace [88]. As discussed by Christina R Victor [89], scholars may also shift from a problem-based focus to preventive aspects, namely toward enhancing healthy social habits. Along these lines, the theoretical framework of the studies might be enriched, going beyond this consideration of loneliness as a contingent and transient or chronic experience by embracing other philosophical perspectives that consider it to be the existential condition of each of us; thus, loneliness is a permanent and unavoidable condition generated by the activity of consciousness [90].

Limitations

This study explicitly concentrated on loneliness and social isolation among older adults. Thus, it contributes most in situations where loneliness and social isolation can be suspected or are already present among older adults, which may be seen as a limitation of this study. In addition, the existing setup could be broadened to include other dimensions related to loneliness and social isolation, and other technologies to yield larger data sets for future reviews. For instance, there are studies using smartphones for detecting and predicting older adults' loneliness and social isolation [91,92]; studies also consider numerous forms of technical systems different from the ones we considered, aiming to reduce older adults' loneliness [93,94]. There are also other studies conducted regarding the effects of loneliness on technology adoption [95], and those neither exploiting physical ICTs [96,97] nor assessing the technology directly in relation to loneliness or social isolation but in terms of related factors [98,99]; these are beyond the scope of this review but closely related to it. However, we did not include all possible technologies and approaches in our study to keep it focused.

Comparison With Prior Work

This review differs from other reviews addressing ICT technologies to assess older adults' loneliness and social isolation focusing on mediating human-human interactions. It addresses ICTs that perform by themselves and do not aim to foster human relationships as the main objective of their interventions. This study also complements some other systematic literature reviews by addressing partially overlapping technologies, namely robots and smart houses; thus, it provides a more comprehensive view of the research in this regard. Further, to our knowledge, this review is the first one compiling the studies on detection and prediction of older adults' loneliness and social isolation, which is an emerging field that has received only little attention thus far [100].

Conclusions

This paper provides a comprehensive overview of the existing attempts aiming to combat older adults' loneliness and social isolation using physical ICTs, namely robots, wearables, and smart homes. Our findings demonstrate that physical ICTs such as smart home solutions can help detect and predict loneliness and social isolation, and technologies such as robotic pets and some other social robots can help alleviate loneliness to some extent. Technologies such as social robots and some of the smart home solutions that can react to human behavior or interact with people are promising, and the literature on these topics has increased recently. These findings have relevant implications to the discussion on combating loneliness and social isolation among older adults that challenge the prejudices about the use of technology in this sensitive area. The results benefit the academic community by accumulating research evidence and finding future research targets, as well as informing other professionals and practitioners of the current state of research, developed solutions, and interventions conducted. The results are also useful in the COVID-19 era, where it is extremely important to find solutions to cope with social isolation and loneliness.

Acknowledgments

This research has received funding from The Intelligent Society Platform (INSO). INSO is an Academy of Finland funded profiling action, operating in Tampere University.

Conflicts of Interest

None declared.

References

1. World Health Organization. World Report on Ageing and Health. Luxembourg: World Health Organization; 2015.
2. Jamieson L, Simpson R. Living Alone: Globalization, Identity and Belonging. London: Palgrave Macmillan; 2013.
3. Böger A, Huxhold O. Age-related changes in emotional qualities of the social network from middle adulthood into old age: How do they relate to the experience of loneliness? Psychol Aging 2018 May;33(3):482-496. [doi: 10.1037/pag0000222] [Medline: 29446969]
4. Huxhold O, Fiori KL, Windsor TD. The dynamic interplay of social network characteristics, subjective well-being, and health: the costs and benefits of socio-emotional selectivity. Psychol Aging 2013 Mar;28(1):3-16. [doi: 10.1037/a0030170] [Medline: 23066804]
5. Wrzus C, Hänel M, Wagner J, Neyer FJ. Social network changes and life events across the life span: a meta-analysis. Psychol Bull 2013 Jan;139(1):53-80. [doi: 10.1037/a0028601] [Medline: 22642230]
6. Luchetti M, Lee JH, Aschwanden D, Sesker A, Strickhouser JE, Terracciano A, et al. The trajectory of loneliness in response to COVID-19. Am Psychol 2020 Oct;75(7):897-908 [FREE Full text] [doi: 10.1037/amp0000690] [Medline: 32567879]
7. van Tilburg TG, Steinmetz S, Stolte E, van der Roest H, de Vries DH. Loneliness and mental health during the COVID-19 pandemic: a study among Dutch older adults. J Gerontol B Psychol Sci Soc Sci 2021 Aug;76(7):e249-e255 [FREE Full text] [doi: 10.1093/geronb/gbaa111] [Medline: 32756931]
8. National Academies of Sciences, Engineering, and Medicine. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington DC: The National Academies Press; Feb 2020.
9. Peplau LA. Loneliness research: basic concepts and findings. In: Inarason IG, Sarason BR, editors. Social Support: Theory, Research and Applications. Dordrecht: Springer; 1985:269-286. [doi: 10.1007/978-94-009-5115-0_15]
10. Perlman D, Peplau LA. Loneliness. In: Friedman H, editor. Encyclopedia of Mental Health, Vol 2. San Diego: Academic Press; 1998:571-581.
11. Weiss R. Loneliness: The Experience of Emotional and Social Isolation. Cambridge, MA: MIT Press; 1973.
12. Perlman D. European and Canadian studies of loneliness among seniors. Can J Aging 2004;23(2):181-188. [doi: 10.1353/cja.2004.0025] [Medline: 15334817]
13. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
14. Beller J, Wagner A. Loneliness, social isolation, their synergistic interaction, and mortality. Health Psychol 2018 Sep;37(9):808-813. [doi: 10.1037/hea0000605] [Medline: 30138019]
15. Hawkley LC, Kocherginsky M. Transitions in loneliness among older adults: a 5-year follow-up in the National Social Life, Health, and Aging Project. Res Aging 2018 Apr;40(4):365-387 [FREE Full text] [doi: 10.1177/0164027517698965] [Medline: 29519211]
16. Pinquart M, Sörensen S. Risk factors for loneliness in adulthoodold age: a meta-analysis. In: Inhohov SP, editor. Advances in Psychology Research, Vol. 19. Hauppauge, NY: Nova Science Publishers; 2003:111-143.
17. Hawkley LC, Kozloski M, Wong J. A Profile of Social Connectedness in Older Adults. 2017. URL: https://connect2affect.org/wp-content/uploads/2020/06/A-Profile-of-Social-Connectedness-1-1.pdf
18. Hansen T, Slagsvold B. Late-life loneliness in 11 European countries: results from the generations and gender survey. Soc Indic Res 2015 Sep;129(1):445-464. [doi: 10.1007/s11205-015-1111-6]
19. Nyqvist F, Nygård M, Scharf T. Loneliness amongst older people in Europe: a comparative study of welfare regimes. Eur J Ageing 2019 Jun;16(2):133-143 [FREE Full text] [doi: 10.1007/s10433-018-0487-y] [Medline: 31139028]
20. Cudjoe T, Roth D, Szanton S, Wolff J, Boyd C, Thorpe R. The epidemiology of social isolation: national health and aging trends study. J Gerontol B Psychol Sci Soc Sci 2020 Jan;75(1):107-113 [FREE Full text] [doi: 10.1093/geronb/gby037] [Medline: 29590462]
21. Holt-Lunstad J. The potential public health relevance of social isolation and loneliness: prevalence, epidemiology, and risk factors. Public Policy & Aging Report 2017;27(4):127-130. [doi: 10.1093/ppar/prx030]
22. Ong AD, Uchino BN, Wethington E. Loneliness and health in older adults: a mini-review and synthesis. Gerontology 2016;62(4):443-449 [FREE Full text] [doi: 10.1159/000441651] [Medline: 26539997]
23. Prohaska T, Burholt V, Burns A, Golden J, Hawkley L, Lawlor B, et al. Consensus statement: loneliness in older adults, the 21st century social determinant of health? BMJ Open 2020 Aug;10(8):e034967 [FREE Full text] [doi: 10.1136/bmjopen-2019-034967] [Medline: 32788184]
24. Barreto M, Victor C, Hammond C, Eccles A, Richins MT, Qualter P. Loneliness around the world: age, gender, and cultural differences in loneliness. Pers Individ Dif 2021 Feb;169:110066 [FREE Full text] [doi: 10.1016/j.paid.2020.110066] [Medline: 33536694]

25. Borys S, Perlman D. Gender differences in loneliness. Pers Soc Psychol Bull 1985 Mar;11(1):63-74. [doi: 10.1177/0146167285111006]
26. Rico-Uribe LA, Caballero FF, Martín-María N, Cabello M, Ayuso-Mateos JL, Miret M. Association of loneliness with all-cause mortality: a meta-analysis. PLoS One 2018;13(1):e0190033 [FREE Full text] [doi: 10.1371/journal.pone.0190033] [Medline: 29300743]
27. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013 Apr;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
28. Valtorta NK, Kanaan M, Gilbody S, Hanratty B. Loneliness, social isolation and risk of cardiovascular disease in the English Longitudinal Study of Ageing. Eur J Prev Cardiol 2018 Sep;25(13):1387-1396. [doi: 10.1177/2047487318792696] [Medline: 30068233]
29. Boss L, Kang D, Branson S. Loneliness and cognitive function in the older adult: a systematic review. Int Psychogeriatr 2015 Apr;27(4):541-553. [doi: 10.1017/S1041610214002749] [Medline: 25554219]
30. Cacioppo JT, Hawkley LC. Perceived social isolation and cognition. Trends Cogn Sci 2009 Oct;13(10):447-454 [FREE Full text] [doi: 10.1016/j.tics.2009.06.005] [Medline: 19726219]
31. Wang J, Mann F, Lloyd-Evans B, Ma R, Johnson S. Associations between loneliness and perceived social support and outcomes of mental health problems: a systematic review. BMC Psychiatry 2018 May;18(1):156 [FREE Full text] [doi: 10.1186/s12888-018-1736-5] [Medline: 29843662]
32. Courtin E, Knapp M. Social isolation, loneliness and health in old age: a scoping review. Health Soc Care Community 2017 May;25(3):799-812. [doi: 10.1111/hsc.12311] [Medline: 26712585]
33. Beutel ME, Klein EM, Brähler E, Reiner I, Jünger C, Michal M, et al. Loneliness in the general population: prevalence, determinants and relations to mental health. BMC Psychiatry 2017 Mar;17(1):97 [FREE Full text] [doi: 10.1186/s12888-017-1262-x] [Medline: 28320380]
34. Richard A, Rohrmann S, Vandeleur CL, Schmid M, Barth J, Eichholzer M. Loneliness is adversely associated with physical and mental health and lifestyle factors: Results from a Swiss national survey. PLoS One 2017 Jul;12(7):e0181442 [FREE Full text] [doi: 10.1371/journal.pone.0181442] [Medline: 28715478]
35. Paul C, Ayis S, Ebrahim S. Psychological distress, loneliness and disability in old age. Psychol Health Med 2006 May;11(2):221-232. [doi: 10.1080/13548500500262945] [Medline: 17129910]
36. Jackson J, Cochran SD. Loneliness and psychological distress. J Psychol 1991 May;125(3):257-262. [doi: 10.1080/00223980.1991.10543289] [Medline: 1880752]
37. Choi H, Irwin MR, Cho HJ. Impact of social isolation on behavioral health in elderly: systematic review. World J Psychiatry 2015 Dec;5(4):432-438 [FREE Full text] [doi: 10.5498/wjp.v5.i4.432] [Medline: 26740935]
38. McHugh JE, Lawlor BA. Perceived stress mediates the relationship between emotional loneliness and sleep quality over time in older adults. Br J Health Psychol 2013 Sep;18(3):546-555. [doi: 10.1111/j.2044-8287.2012.02101.x] [Medline: 22988915]
39. Pels F, Kleinert J. Loneliness and physical activity: a systematic review. Int Rev Sport Exerc Psychol 2016 Jun;9(1):231-260. [doi: 10.1080/1750984X.2016.1177849]
40. Cornwell EY, Waite LJ. Social disconnectedness, perceived isolation, and health among older adults. J Health Soc Behav 2009 Mar;50(1):31-48 [FREE Full text] [doi: 10.1177/002214650905000103] [Medline: 19413133]
41. Hawton A, Green C, Dickens AP, Richards SH, Taylor RS, Edwards R, et al. The impact of social isolation on the health status and health-related quality of life of older people. Qual Life Res 2011 Feb;20(1):57-67. [doi: 10.1007/s11136-010-9717-2] [Medline: 20658322]
42. Leigh-Hunt N, Bagguley D, Bash K, Turner V, Turnbull S, Valtorta N, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
43. Fratiglioni L, Wang H, Ericsson K, Maytan M, Winblad B. Influence of social network on occurrence of dementia: a community-based longitudinal study. Lancet 2000 Apr;355(9212):1315-1319. [doi: 10.1016/S0140-6736(00)02113-9] [Medline: 10776744]
44. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: a systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
45. Hagan R, Manktelow R, Taylor BJ, Mallett J. Reducing loneliness amongst older people: a systematic search and narrative review. Aging Ment Health 2014 Jan;18(6):683-693. [doi: 10.1080/13607863.2013.875122] [Medline: 24437736]
46. Poscia A, Stojanovic J, La Milia DI, Duplaga M, Grysztar M, Moscato U, et al. Interventions targeting loneliness and social isolation among the older people: an update systematic review. Exp Gerontol 2018 Feb;102:133-144. [doi: 10.1016/j.exger.2017.11.017] [Medline: 29199121]
47. Shelton BE, Uz C. Immersive technology and the elderly: a mini-review. Gerontology 2015 Feb;61(2):175-185 [FREE Full text] [doi: 10.1159/000365754] [Medline: 25502624]

48. Vichitvanichphong S, Talaei-Khoei A, Kerr D, Ghapanchi AH. Adoption of assistive technologies for aged care: a realist review of recent studies. : IEEE; 2014 Presented at: 2014 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI, USA p. 2706-2715.. [doi: 10.1109/hicss.2014.341]
49. Duffy BR. Anthropomorphism and the social robot. Rob Auton Syst 2003 Mar;42(3-4):177-190. [doi: 10.1016/s0921-8890(02)00374-3]
50. Fong T, Nourbakhsh I, Dautenhahn K. A survey of socially interactive robots. Rob Auton Syst 2003 Mar;42(3-4):143-166. [doi: 10.1016/s0921-8890(02)00372-x]
51. Chan M, Campo E, Estève D, Fourniols J. Smart homes—current features and future perspectives. Maturitas 2009 Oct;64(2):90-97. [doi: 10.1016/j.maturitas.2009.07.014] [Medline: 19729255]
52. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]
53. Shishehgar M, Kerr D, Blake J. A systematic review of research into how robotic technology can help older people. Smart Health 2018 Jun;7-8:1-18. [doi: 10.1016/j.smhl.2018.03.002]
54. Pal D, Triyanson T, Funilkul S. Smart homes and quality of life for the elderly: a systematic review. Presented at: 2017 IEEE International Symposium on Multimedia (ISM); 2017 Dec 11-13; Taichung, Taiwan p. 413-419.. [doi: 10.1109/ism.2017.83]
55. Baecker AN, Geiskkovitch DY, González AL, Young JE. Emotional support domestic robots for healthy older adults: conversational prototypes to help with loneliness. New York, NY, USA.: Association for Computing Machinery; 2020 Presented at: HRI '20: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction; March 23-26, 2020; Cambridge, United Kingdom p. 122-124.. [doi: 10.1145/3371382.3378279]
56. Gasteiger N, Loveys K, Law M, Broadbent E. Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. Clin Interv Aging 2021 May;16:941-971 [FREE Full text] [doi: 10.2147/CIA.S282709] [Medline: 34079242]
57. Isabet B, Pino M, Lewis M, Benveniste S, Rigaud A. Social telepresence robots: a narrative review of experiments involving older adults before and during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Mar;18(7):3597 [FREE Full text] [doi: 10.3390/ijerph18073597] [Medline: 33808457]
58. Jecker NS. You've got a friend in me: sociable robots for older adults in an age of global pandemics. Ethics Inf Technol 2020 Jul:1-9 [FREE Full text] [doi: 10.1007/s10676-020-09546-y] [Medline: 32837286]
59. Marques G. Ambient assisted living and internet of things. In: Cardoso P, Monteiro J, Semiao J, Rodrigues J, editors. Harnessing the Internet of Everything (IoE) for Accelerated Innovation Opportunities. Hershey, PA, USA: IGI Global; 2019:100-115.
60. Demiris G. Smart homes and ambient assisted living in an aging society. New opportunities and challenges for biomedical informatics. Methods Inf Med 2018 Jan;47(1):56-57. [doi: 10.1055/s-0038-1625127] [Medline: 18213428]
61. 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;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
62. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A smart-home system to unobtrusively and continuously assess loneliness in older adults. IEEE J Transl Eng Health Med 2016 Jul;4:2800311 [FREE Full text] [doi: 10.1109/JTEHM.2016.2579638] [Medline: 27574577]
63. Goonawardene N, Toh X, Tan H. Sensor-driven detection of social isolation in community-dwelling elderly. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Applications, Services and Contexts ITAP 2017. Lecture Notes in Computer Science, vol 10298. Cham: Springer; 2017:378-392.
64. Petersen J, Austin D, Kaye JA, Pavel M, Hayes TL. Unobtrusive in-home detection of time spent out-of-home with applications to loneliness and physical activity. IEEE J Biomed Health Inform 2014 Sep;18(5):1590-1596 [FREE Full text] [doi: 10.1109/JBHI.2013.2294276] [Medline: 25192570]
65. Petersen J, Austin D, Mattek N, Kaye J. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: a longitudinal mixed effects model. PLoS One 2015 Oct;10(10):e0139643 [FREE Full text] [doi: 10.1371/journal.pone.0139643] [Medline: 26437228]
66. Walsh L, Kealy A, Loane J, Doyle J, Bond R. Inferring health metrics from ambient smart home data. : IEEE; 2014 Presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 2-5, 2014; Belfast, UK p. 27-32. [doi: 10.1109/BIBM.2014.6999237]
67. Huynh S, Tan H, Lee Y. Towards unobtrusive mental well-being monitoring for independent-living elderly. In: Proceedings of the 4th International on Workshop on Physical Analytics. New York, NY: Association for Computing Machinery; 2017 Presented at: WPA '17; June 19, 2017; Niagara Falls, New York, USA p. 1-6. [doi: 10.1145/3092305.3092311]
68. Martinez A, Ortiz V, Estrada H, Gonzalez M. A predictive model for automatic detection of social isolation in older adults. : IEEE; 2017 Presented at: 2017 International Conference on Intelligent Environments (IE); August 21-25, 2017; Seoul, Korea (South) p. 68-75. [doi: 10.1109/IE.2017.35]

69. Banks MR, Willoughby LM, Banks WA. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. J Am Med Dir Assoc 2008 Mar;9(3):173-177. [doi: 10.1016/j.jamda.2007.11.007] [Medline: 18294600]
70. Casey D, Barrett E, Kovacic T, Sancarlo D, Ricciardi F, Murphy K, et al. The perceptions of people with dementia and key stakeholders regarding the use and impact of the social robot MARIO. Int J Environ Res Public Health 2020 Nov;17(22):8621 [FREE Full text] [doi: 10.3390/ijerph17228621] [Medline: 33233605]
71. Chen S, Moyle W, Jones C, Petsky H. A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int Psychogeriatr 2020 Aug;32(8):981-991. [doi: 10.1017/S1041610220000459] [Medline: 32284080]
72. Fields N, Xu L, Greer J, Murphy E. Shall I compare thee…to a robot? An exploratory pilot study using participatory arts and social robotics to improve psychological well-being in later life. Aging Ment Health 2021 Mar;25(3):575-584. [doi: 10.1080/13607863.2019.1699016] [Medline: 31851830]
73. Follmann A, Schollemann F, Arnolds A, Weismann P, Laurentius T, Rossaint R, et al. Reducing loneliness in stationary geriatric care with robots and virtual encounters—a contribution to the COVID-19 pandemic. Int J Environ Res Public Health 2021 May;18(9):4846 [FREE Full text] [doi: 10.3390/ijerph18094846] [Medline: 34062776]
74. Hudson J, Ungar R, Albright L, Tkatch R, Schaeffer J, Wicker E. Robotic pet use among community-dwelling older adults. J Gerontol B Psychol Sci Soc Sci 2020 Oct;75(9):2018-2028 [FREE Full text] [doi: 10.1093/geronb/gbaa119] [Medline: 32789476]
75. Lazar A, Thompson H, Piper A, Demiris G. Rethinking the design of robotic pets for older adults. In: Proceedings of the 2016 ACM Conference on Designing Interactive Systems. New York, NY, USA: Association for Computing Machinery; 2016 Presented at: DIS '16: Designing Interactive Systems Conference 2016; June 4-8, 2016; Brisbane QLD Australia p. 1034-1046. [doi: 10.1145/2901790.2901811]
76. Papadopoulos C, Castro N, Nigath A, Davidson R, Faulkes N, Menicatti R, et al. The CARESSES randomised controlled trial: exploring the health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. Int J Soc Robot 2021 Apr:1-12 [FREE Full text] [doi: 10.1007/s12369-021-00781-x] [Medline: 33907589]
77. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
78. Sidner CL, Bickmore T, Nooraie B, Rich C, Ring L, Shayganfar M, et al. Creating new technologies for companionable agents to support isolated older adults. ACM Trans Interact Intell Syst 2018 Aug;8(3):1-27. [doi: 10.1145/3213050]
79. Tkatch R, Wu L, MacLeod S, Ungar R, Albright L, Russell D, et al. Reducing loneliness and improving well-being among older adults with animatronic pets. Aging Ment Health 2021 Jul;25(7):1239-1245. [doi: 10.1080/13607863.2020.1758906] [Medline: 32363903]
80. Curumsing MK, Fernando N, Abdelrazek M, Vasa R, Mouzakis K, Grundy J. Emotion-oriented requirements engineering: a case study in developing a smart home system for the elderly. J Syst Softw 2019 Jan;147:215-229. [doi: 10.1016/j.jss.2018.06.077]
81. Appel L, Appel E, Bogler O, Wiseman M, Cohen L, Ein N, et al. Older adults with cognitive and/or physical impairments can benefit from immersive virtual reality experiences: a feasibility study. Front Med (Lausanne) 2020 Jan;6:329 [FREE Full text] [doi: 10.3389/fmed.2019.00329] [Medline: 32010701]
82. Lin C, Lee C, Lally D, Coughlin J. Impact of virtual reality (VR) experience on older adults' well-being. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Applications in Health, Assistance, and Entertainment. ITAP 2018. Lecture Notes in Computer Science, vol 10927. Cham: Springer; 2018:89-100.
83. Brandenburgh A, van Breda W, van der Ham W, Klein M, Moeskops L, Roelofsma P. Increasing physical and social activity through virtual coaching in an ambient environment. In: Ślezak D, Schaefer G, Vuong ST, Kim YS, editors. Active Media Technology. AMT 2014. Lecture Notes in Computer Science, vol 8610. Cham: Springer; 2014:487-500.
84. Baisch S, Kolling T, Schall A, Rühl S, Selic S, Kim Z, et al. Acceptance of social robots by elder people: Does psychosocial functioning matter? Int J of Soc Robotics 2017 Jan;9(2):293-307. [doi: 10.1007/s12369-016-0392-5]
85. Robinson H, Broadbent E, MacDonald B. Group sessions with Paro in a nursing home: structure, observations and interviews. Australas J Ageing 2016 Jun;35(2):106-112. [doi: 10.1111/ajag.12199] [Medline: 26059390]
86. Pérez-Gómez A. Attunement: Architectural Meaning after the Crisis of Modern Science. Cambridge, Massachusetts: The MIT Press; 2016.
87. Vesely D. Architecture as a humanistic discipline. In: Bandyopadhyay S, Lomholt J, Temple N, Tobe R, editors. The Humanities in Architectural Design. A Contemporary and Historical Perspective. Oxfordshire, England, UK: Routledge; 2010:211-222.
88. International Telecommunication Union. Digital trends in Europe 2021 - ICT trends and developments in Europe, 2017-2020.: ITU Publications; 2021. URL: http://handle.itu.int/11.1002/pub/817b8884-en [accessed 2021-06-06]
89. Victor CR. The languages of loneliness: developing a vocabulary for researching social health. In: Nieto F, Rubio R, editors. Loneliness and the Built Environment. Datutop Occasional Papers 40. Tampere: Tampere University School of Architecture; 2021:51-71.

90. Mijuskovic BL. Theories of consciousness and loneliness. In: Nieto F, Rubio R, editors. Loneliness and the Built Environment. Datutop Occasional Papers 40. Tampere: Tampere University School of Architecture; 2021:19-49.
91. Campos W, Martinez A, Sanchez W, Estrada H, Favela J, Perez J. Inferring social isolation in older adults through ambient intelligence and social networking sites. CyS 2016 Mar;20(1):143-152. [doi: 10.13053/cys-20-1-2193]
92. Sanchez W, Martinez A, Campos W, Estrada H, Pelechano V. Inferring loneliness levels in older adults from smartphones. J Ambient Intell Smart Environ 2015 Jan;7(1):85-98. [doi: 10.3233/AIS-140297]
93. Goumopoulos C, Papa I, Stavrianos A. Development and evaluation of a mobile application suite for enhancing the social inclusion and well-being of seniors. Informatics 2017 Jun;4(3):15. [doi: 10.3390/informatics4030015]
94. van der Heide LA, Willems CG, Spreeuwenberg MD, Rietman J, de Witte LP. Implementation of CareTV in care for the elderly: the effects on feelings of loneliness and safety and future challenges. TAD 2012 Jan;24(4):283-291. [doi: 10.3233/tad-120359]
95. Garattini C, Wherton J, Prendergast D. Linking the lonely: an exploration of a communication technology designed to support social interaction among older adults. Univ Access Inf Soc 2011 Jun;11(2):211-222. [doi: 10.1007/s10209-011-0235-y]
96. Wu Y, Cristancho-Lacroix V, Fassert C, Faucounau V, de Rotrou J, Rigaud A. The attitudes and perceptions of older adults with mild cognitive impairment toward an assistive robot. J Appl Gerontol 2016 Jan;35(1):3-17. [doi: 10.1177/0733464813515092] [Medline: 24652924]
97. Karim H, Lokman A, Redzuan F. Older adults perspective and emotional respond on robot interaction. Presented at: 2016 4th International Conference on User Science and Engineering (i-USEr); August 23-25, 2016; Melaka, Malaysia p. 95-99. [doi: 10.1109/IUSER.2016.7857941]
98. Wu Y, Wrobel J, Cristancho-Lacroix V, Kamali L, Chetouani M, Duhaut D, et al. Designing an assistive robot for older adults: the ROBADOM project. IRBM 2013 Apr;34(2):119-123. [doi: 10.1016/j.irbm.2013.01.003]
99. Bellini G, Cipriano M, Comai S, De Angeli N, Gargano J, Gianella M, et al. Understanding social behaviour in a health-care facility from localization data: a case study. Sensors (Basel) 2021 Mar;21(6):2147 [FREE Full text] [doi: 10.3390/s21062147] [Medline: 33803913]
100. Campos W, Martinez A, Sanchez W, Estrada H, Castro-Sánchez NA, Mujica D. A systematic review of proposals for the social integration of elderly people using ambient intelligence and social networking sites. Cogn Comput 2016 Feb;8(3):529-542. [doi: 10.1007/s12559-016-9382-z]

Abbreviations

AAL: ambient assisted living
ICT: information and communication technology
UCLA: University of California Los Angeles
VR: virtual reality

Edited by R Kukafka; submitted 20.02.21; peer-reviewed by J Taylor, M Jordan-Marsh; comments to author 10.05.21; revised version received 30.06.21; accepted 08.11.21; published 30.12.21.

Please cite as:
Latikka R, Rubio-Hernández R, Lohan ES, Rantala J, Nieto Fernández F, Laitinen A, Oksanen A
Older Adults' Loneliness, Social Isolation, and Physical Information and Communication Technology in the Era of Ambient Assisted Living: A Systematic Literature Review
J Med Internet Res 2021;23(12):e28022
URL: https://www.jmir.org/2021/12/e28022
doi:10.2196/28022
PMID:34967760

Viewpoint

Open-source Software Sustainability Models: Initial White Paper From the Informatics Technology for Cancer Research Sustainability and Industry Partnership Working Group

Ye Ye[1], BMED, MSPH, PhD; Seemran Barapatre[1]; Michael K Davis[1], MS; Keith O Elliston[2,3,4], PhD; Christos Davatzikos[5], PhD; Andrey Fedorov[6], PhD; Jean-Christophe Fillion-Robin[7], MSc; Ian Foster[8], PhD; John R Gilbertson[1], MD; Andras Lasso[9], PhD; James V Miller[10], PhD; Martin Morgan[11], PhD; Steve Pieper[12], PhD; Brigitte E Raumann[13], PhD; Brion D Sarachan[10], MEng; Guergana Savova[14], PhD; Jonathan C Silverstein[1], MD, MS; Donald P Taylor[1], MBA, PhD, CLP; Joyce B Zelnis[1], BS; Guo-Qiang Zhang[15], MS, PhD; Jamie Cuticchia[2], PhD, JD; Michael J Becich[1], MD, PhD

[1]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[2]Axiomedix, Inc., Bedford, MA, United States
[3]PHEMI Systems Corp., Vancouver, BC, Canada
[4]tranSMART foundation, Wakefield, MA, United States
[5]Department of Radiology, School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[6]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[7]Kitware Inc., Clifton Park, NY, United States
[8]Department of Computer Science, University of Chicago, Chicago, IL, United States
[9]The Perk Lab for Percutaneous Surgery, School of Computing, Queen's University, Kingston, ON, Canada
[10]GE Global Research, Niskayuna, NY, United States
[11]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, United States
[12]Isomics Inc., Cambridge, MA, United States
[13]Globus, University of Chicago, Chicago, IL, United States
[14]Boston Children's Hospital, Harvard Medical School, Boston, MA, United States
[15]The University of Texas Health Science Center at Houston, Houston, TX, United States

Corresponding Author:
Michael J Becich, MD, PhD
Department of Biomedical Informatics
School of Medicine
University of Pittsburgh
5607 Baum Blvd Office 502, 15206
Pittsburgh, PA
United States
Phone: 1 412 648 9244
Email: becich@pitt.edu

Abstract

Background: The National Cancer Institute Informatics Technology for Cancer Research (ITCR) program provides a series of funding mechanisms to create an ecosystem of open-source software (OSS) that serves the needs of cancer research. As the ITCR ecosystem substantially grows, it faces the challenge of the long-term sustainability of the software being developed by ITCR grantees. To address this challenge, the ITCR sustainability and industry partnership working group (SIP-WG) was convened in 2019.

Objective: The charter of the SIP-WG is to investigate options to enhance the long-term sustainability of the OSS being developed by ITCR, in part by developing a collection of business model archetypes that can serve as sustainability plans for ITCR OSS development initiatives. The working group assembled models from the ITCR program, from other studies, and from the engagement

of its extensive network of relationships with other organizations (eg, Chan Zuckerberg Initiative, Open Source Initiative, and Software Sustainability Institute) in support of this objective.

Methods: This paper reviews the existing sustainability models and describes 10 OSS use cases disseminated by the SIP-WG and others, including 3D Slicer, Bioconductor, Cytoscape, Globus, i2b2 (Informatics for Integrating Biology and the Bedside) and tranSMART, Insight Toolkit, Linux, Observational Health Data Sciences and Informatics tools, R, and REDCap (Research Electronic Data Capture), in 10 sustainability aspects: governance, documentation, code quality, support, ecosystem collaboration, security, legal, finance, marketing, and dependency hygiene.

Results: Information available to the public reveals that all 10 OSS have effective governance, comprehensive documentation, high code quality, reliable dependency hygiene, strong user and developer support, and active marketing. These OSS include a variety of licensing models (eg, general public license version 2, general public license version 3, Berkeley Software Distribution, and Apache 3) and financial models (eg, federal research funding, industry and membership support, and commercial support). However, detailed information on ecosystem collaboration and security is not publicly provided by most OSS.

Conclusions: We recommend 6 essential attributes for research software: alignment with unmet scientific needs, a dedicated development team, a vibrant user community, a feasible licensing model, a sustainable financial model, and effective product management. We also stress important actions to be considered in future ITCR activities that involve the discussion of the sustainability and licensing models for ITCR OSS, the establishment of a central library, the allocation of consulting resources to code quality control, ecosystem collaboration, security, and dependency hygiene.

(J Med Internet Res 2021;23(12):e20028) doi:10.2196/20028

KEYWORDS

open-source software; sustainability; licensing model; financial model; product management; cancer informatics

Introduction

Background

The Informatics Technology for Cancer Research (ITCR) program [1] was established by the National Cancer Institute (NCI) in 2012 to create an ecosystem of open-source software (OSS) that serves the needs of cancer research. ITCR supports informatics technology development initiated by cancer research investigators and includes 4 extramural divisions: cancer biology, cancer control and population science, cancer prevention, and cancer treatment and diagnosis. The coordinating body for ITCR is the NCI Center for Biomedical Informatics and Informatics Technology.

The specific goals of ITCR include (1) promoting the integration of informatics technology development with hypothesis-driven cancer research and translational or clinical investigations; (2) providing flexible, scalable, and sustainable support using multiple mechanisms matched to the various needs and different stages of informatics technology development throughout the development life cycle; (3) promoting interdisciplinary collaboration and public–private partnerships in technology development and distribution; (4) promoting data sharing and development of informatics tools to enable data sharing; (5) promoting technology dissemination and software reuse; (6) promoting communication and interaction among development teams; and (7) leveraging the NCI program expertise and resources across the institute and bridging gaps in the existing NCI grant portfolios for informatics.

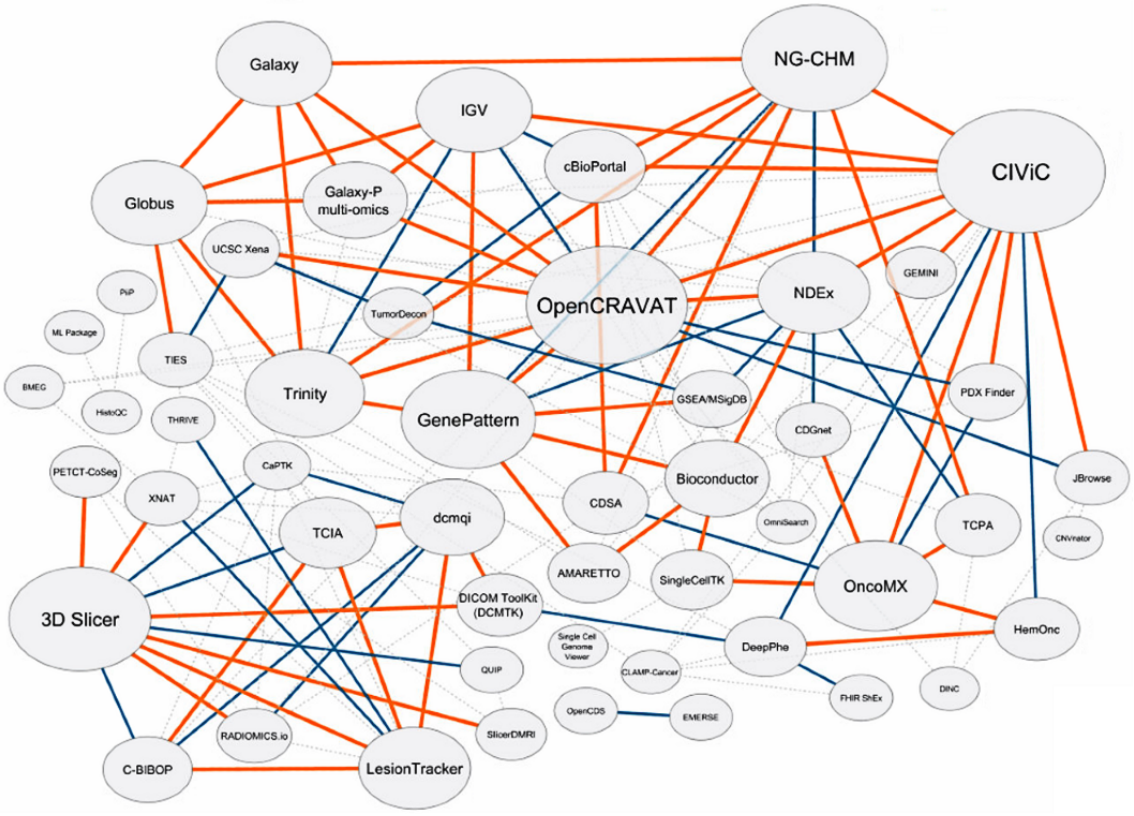
The scope of the ITCR program is to serve informatics needs that span the cancer research continuum. The ITCR program provides a series of funding mechanisms that support informatics resources across the development life cycle, including the creation of innovative methods and algorithms (R21), early-stage software development (R21), advanced stage software development (U24), and the sustainment of high-value resources (U24) on which the cancer research and translational informatics community has come to depend (Table 1). The program also offers supplements (*competitive revisions*) to currently funded NCI grantees to incorporate ITCR technologies into their ongoing research. Current funding opportunities are available on the ITCR website [2].

Table 1. Informatics Technology for Cancer Research (ITCR) funding mechanisms.

Mechanism	Purpose	Awards before September 9, 2020	Direct cost cap
R21	Innovative informatics methods and algorithms	25	US $275,000 over 2 years
U01	Early-stage software development	34	US $300,000 per year for up to 3 years
U24	Advanced stage software development	40	US $600,000 per year for up to 5 years
U24	Sustainment of high-value resources	6	No budget cap and up to 5 years of support
Competitive revisions (new)	Adoption, adaptation, and integration of ITCR tools and resources	1	US $100,000 per year for up to 2 years

This series of funding mechanisms is innovative and unique across all National Institutes of Health (NIH) institutes and centers. These mechanisms address a fundamental need to create a computational infrastructure that is interoperable and collaborative, linking many informatics and computational biology teams performing translational informatics. The ITCR ecosystem has grown substantially and now includes 55 funded efforts that are highly collaborative, as evidenced by its *connectivity map* (Figure 1). This map is copied from the Network Data Exchange website [3,4]. In this map, each node represents a project funded under ITCR. Links among these nodes represent connections among projects. Existing connections are represented by orange solid lines, ongoing connections by blue solid lines, and proposed connections by gray dashed lines. The node size is determined by the connectivity score, which is calculated by assigning 0 points for each proposed connection, 1 point for each ongoing connection, and 3 points for each existing connection. A large node usually indicates that the project has many existing connections with other projects. The connectivity scores are available on the Network Data Exchange website.

Figure 1. The map of Informatics Technology for Cancer Research projects.



As the ITCR program moves into its second phase, it faces the challenge of long-term sustainability for the software being developed by its grantees. Whether viewed from the angle of a single funded project or all ITCR-funded projects, some of the software will naturally *graduate* upon reaching maturity to leave room for continuing innovation through the program. As mature projects often lead to complex and successful products based on years of investment in human effort, funding, and cumulative expertise, these projects need to move into the next phase of support rather than risk being abandoned.

Addressing the challenge of the projects' long-term sustainability was the primary task of the ITCR sustainability and industry partnership working group (SIP-WG) [5], which was convened in 2019. The working group initially set the goals of addressing 4 topics of interest to the translational cancer informatics community: (1) to publish a collection of case studies of successfully disseminated software products supported by open-source licenses and to provide practical examples of approaches that have proven viable for licensing and sustainability, (2) to develop a workflow or decision tree to support informed decision-making consistent with ITCR expectations and the future licensing needs of open-source tools, (3) to provide a licensing consultancy service in collaboration with the ITCR program, and (4) to develop a collection of business model archetypes that can serve as starting templates and to formally document the dissemination and sustainability plans for new software development initiatives. The ITCR licensing resources will represent *best practice* approaches and leverage our extensive network of relationships with organizations such as the Open Source Initiative, the Software Sustainability Institute, and the Chan Zuckerberg Initiative to maintain relevant knowledge in this field. As described above, the first major topic—publication of case studies—is the subject of this paper. The remaining 3 topics will be the focus of future white papers and manuscripts by the ITCR SIP-WG.

Literature Review

We briefly introduce several software sustainability models that are present in the literature [6-8]. First, Aartsen et al [8] described 2 models for sustaining digital assets from

public–private partnerships in medical research: the not-for-profit organization model and the distributed network model. The *not-for-profit organization model* uses, for example, a foundation (also discussed by Kuchinke et al [6]) as the backbone organization to assure the maximum value of the assets. The Apache Software Foundation is one such example. An advantage of nonprofits is that they can take a long-term view. The sustainability of nonprofits can be mitigated through memberships. The concept of a foundation has the advantage that the development of an artifact is strongly influenced by academic users, so its design can be focused on scientific goals instead of commercial ones. The disadvantage of the not-for-profit organization model is its dependency on one organization for all digital assets. The *distributed network model* is built on the premise that individual partners who contribute to the development of digital assets have a stake in seeing these assets sustained and gaining future value through further development. The disadvantage of the distributed network model lies in the conflicting missions of research and industry; organizations with a research mission do not focus on producing digital assets that are ready to be commercialized by industry.

Gabella et al [9] provided a comprehensive review that adds 10 models for the sustainability of assets, including 4 noncommercial and 6 commercial models. As a noncommercial model, the *national funding model* supports infrastructure directly through noncyclical funding programs. On the other hand, in the *infrastructure model*, funding agencies set aside a fixed percentage of their research grant volumes to be redistributed among core data resources according to well-defined selection criteria. In the *institutional support model*, funds are provided internally from the institution, whereas the *donation model* depends on external philanthropic funding. In terms of the 6 commercial models, the *content licensing or industrial support model* requires commercial users to pay a fee for access and for-profit use, with the assets being free for noncommercial users (also discussed by Kuchinke et al [6]). The *user subscription model* (also discussed by Chang et al [7]) relies on a subscription for a set period. The *freemium model* (also discussed in Chang et al [7]) provides a core that is free, with add-ons requiring a fee. The *razor and blades model* (also discussed in a Wikipedia introduction for business models of OSS [10] as a commercial model) offers a free initial trial (*razor*) that encourages the continuing future purchases of follow-up services (*blades*). The *mixed model* relies on multiple diversified funding streams. For instance, a common mixed model practice is the combination of OSS with services (provided by companies) on installation, configuration, and troubleshooting. Linux is a familiar example of this. However, the Linux model relies on a large user base, which may not necessarily be the case with biomedical research tools.

In addition to the models discussed in last paragraph, the *macro research and development infrastructure* is based on funding that comes from governmental research grants or from research grants from local or international partner institutes [7]. The *split licensing model* offers a free version under a general public license (GPL) and a commercial version with its own license that does not allow software redistribution (eg, MySQL [Sun Microsystems, Inc] and openClinica [OpenClinica, LLC]) [10].

The current literature has also discussed the importance of the strength and health of the community behind a software product [11-13]. Iaffaldano et al [11] used the sleep stage metaphor to describe developer cycles: the awake stage is when developers are active in the project, the sleep stage is when developers pause their package commit activity, and the dead stage is when developers abandon the project. They further explored the reasons for the stage transitions, listing both personal factors (eg, life event, financial, and change of interest) and project factors (eg, social, changes in the project, and role change) as playing a role. Atiq et al [12] suggested sponsoring of open-source projects in various ways as an increasing number of proprietary firms participate in, sponsor, and offer their developers for open-source projects. Jiménez et al [13] provided 4 recommendations for a sustainable open-source project: (1) making the source code publicly accessible from day 1, (2) making the software easy to discover by providing software metadata via a popular community registry, (3) adopting a licensing system that complies with the licenses of third-party dependencies, and (4) defining clear and transparent contributions, governance, and communication processes. Nyman et al [14,15] discussed code forking (implementing an existing code base found in a separate project) within the context of OSS. The right to fork code is built into the very definition of open source. Code forking can revive community interest in a project or provide an alternative to acquisitions, which was the case with MySQL after Oracle's acquisition of Sun Microsystems. The MySQL code was forked under a different name, MariaDB, because of concerns regarding the governance and future openness of the MySQL code. Nyman and Lindman [14] state, "Given that forking ensures that any project can continue as long as there is sufficient community interest, we have previously described forking as the 'invisible hand of sustainability' in open source." For specifically big biology, Prins et al [16] described the challenges of creating sustainable software solutions: most OSS are developed as prototype software, many OSS are not scaled to terabytes of data, and there is a lack of scientific attribution for software development.

Methods

We conducted a survey among the members of the working group to select a collection of case examples of successfully disseminated software products. We asked each member to provide the *best 3* examples of sustainable OSS to serve as models for ITCR open-source projects. The survey was completed by 13 participants, most of whom were authors of this white paper and had years of experience developing OSS for cancer research. To profile the models of success in sustainability, 22 OSS use cases were provided by this survey, and the top 10 tools were then assigned to authors who were then asked to profile the following models: 3D Slicer [17], Bioconductor [18], Cytoscape [19], Globus [20], i2b2 (Informatics for Integrating Biology and the Bedside) [21] and tranSMART [22], Insight Toolkit (ITK) [23], Linux [24], Observational Health Data Sciences and Informatics (OHDSI) [25], R [26], and REDCap (Research Electronic Data Capture; provides a nonprofit end user license agreement but its code base is not open to individual developers) [27].

After reviewing the literature and discussing it in the ITCR working group, we determined that each OSS use case should be profiled according to recommendations by Nesbitt [28] in his paper: "What does a sustainable open source project look like?" Accordingly, each of the top 10 OSS use cases was profiled in the following aspects of sustainability: governance, documentation, code quality, support, ecosystem collaboration, security, legal issues, financing, marketing, and dependency hygiene. Profiling mainly relies on information that is publicly available. As some of the coauthors are key developers of 3D Slicer (AF, SP, JCFR, JVM, and AL) and Globus (IF and BER), we were able to provide more firsthand information on these 2 cases.

Results

In this section, we examine each OSS use case in terms of these 10 sustainability aspects. Full descriptions of the OSS use cases are available in Multimedia Appendix 1.

Governance

All 10 OSS use cases have a management committee and a technology development team. ITK and REDCap have established consortiums. The 3 models (i2b2 tranSMART, R, and Linux) have established foundations. Stakeholders usually choose a consortium management model during the early stages of software development. In a consortium model, members have stronger control over the direction of development. A consortium management model may later migrate into a foundation model. In a foundation model, the organization considers the interests of all stakeholders, encouraging more new contributors and users to participate in the software development testing process. As a result, foundations usually require serious community efforts and diverse skills (eg, fundraising) [29].

The 6 OSS tools have provided their roadmaps publicly. The i2b2 tranSMART Foundation [30] defines a road map guiding the integration of tranSMART with i2b2 [31]. The 3D Slicer's road map [32] lists community suggestions related to a transition plan for Slicer 4.10 and the proposed changes for Slicer 5.x. Cytoscape's road map [33] shows that it is going down a number of roads simultaneously, including Cytoscape Desktop, Cytoscape Expansion to the Cloud, and Cytoscape Community Outreach. Globus's product road map [34] has plans to provide *research information technology as a service*. ITK's team has been continuously updating its road map [35-37] based on feedback from its community of users and developers as well as from the medical research community. OHDSI has several roadmaps, including an architecture road map [38], a road map for CDM v6.0 [39], and a road map for webAPI [40]. On the other hand, LWN.net, a computing webzine on software for Linux and other Unix-like operating systems, points out that the free software development model is resistant to central planning in general [41]. Although not always reliable, Linux's future can be reasonably predicted by looking at its current projects.

Regular meetings allow stakeholders to make operational decisions and set development priorities. The 3D Slicer's core developers and users meet in person twice a year, and Globus has an annual conference for its users and subscribers. The subgroups usually have more frequent regular meetings. On the other hand, the technical advisory board of Bioconductor meets monthly to develop strategies that ensure the long-term technical suitability of the core infrastructure. To reach a broader group of potential developers and users, some models (3D Slicer and i2b2 tranSMART) provide completely open communication channels, such as web-based forums and recorded webinars.

Owing to the limited amount of public information on these 10 OSS use cases, we do not know the exact size of each core development team or the individual assignments on core infrastructure. If there is a single person handling the complicated details of a critical component, an OSS project will go adrift quickly after losing that key person.

Documentation

All 10 OSS use cases provide documentation to users in various formats, such as user guidebooks (ITK [42], Linux [43], R [44], and 3D Slicer [45]), Wiki pages (3D Slicer [46] and i2b2 [47]), tutorials (Bioconductor [48], Globus [49], Cytoscape [50], tranSMART [51], 3D Slicer [52], and OHDSI [53]), and YouTube (Google, Inc) videos (REDCap [54], 3D Slicer [55], and Cytoscape [56]).

Further documentation is provided to new developers to encourage new contributions to OSS extensions. Bioconductor offers 3 levels of documentation—workflows, package vignettes, and function manual pages [57]—to encourage users to become developers who can make their own algorithms and approaches available to others. Similarly, the Cytoscape *App Ladder* teaches essential skills in app development [58]. R provides a variety of fully developed documentation, adequately covering 2 types of development: writing R extensions and developing R itself (by providing internal structure and coding standards) [59].

Code Quality

Releasing software without testing could be very dangerous to its reliability and reproducibility, so rigorous tests are critical for OSS. Before propagating the latest packages to user-facing repositories, Bioconductor developers conduct tests to ensure overall package integrity and integration with current versions of package dependencies. The 3D Slicer has established infrastructure to continuously run approximately 700 tests for its core application, with the test results being publicly available [60]. However, the quality control of some of the extensions of the 3D Slicer is slightly weaker than that of the core application. The extension contributors themselves manage the code quality and tests, and the 3D Slicer's core developer team does not enforce or verify these extensions. Cytoscape developers use Jenkins to build software projects continuously and test packages thoroughly before releasing them. Globus uses a continuous integration environment, automated tests, multiple prerelease environments, and documented, standardized, human quality assurance testing to ensure code quality, with at least one engineer other than the code author reviewing the code before releasing it to production. Both i2b2 and tranSMART have extensive automated and manual testing as part of their well-defined release processes. ITK had automated nightly builds and tests as far back as 1999, being an early adopter of

this software engineering best practice before the widespread adoption of continuous integration and GitHub (GitHub, Inc). R provides extensive support to facilitate external developers' package testing and release, which includes release guidelines, software packages, and servers for testing [61]. A few models (3D Slicer [62], ITK, and R [63]) enforce a consistent coding style.

Support

All OSS use cases provide support to users and new developers. For example, the OHDSI community provides 2 support channels: the community-based discourse forum provides support for implementing OHDSI tools, proposing or participating in network research studies, and requesting information on OHDSI-related topics [64]; and the GitHub project sites of OHDSI manage specific technical questions through tickets that anyone can issue [65]. Globus has several support options: web-based self-help tools, listserv groups, and a ticketing submission system with a responsive support team. R mainly relies on web-based self-help tools, frequently asked question listings, and subscription-based email lists, including a general R help email list, an R developer list, and an R package developer list. Although these models provide various support channels, Linux and Cytoscape mainly rely on dedicated channels (Linux: LF JIRA [66]; Cytoscape: a specific help desk [67]).

Not all support models for OSS are free. For example, ITK has a 3-way support: (1) ITK's discourse forum enables discussion and mutual help among users, and dedicated volunteers usually provide detailed example codes [68]; (2) the NIH has continued to provide maintenance contracts for bug fixes, incremental improvements, and a moderate level of user support (maintenance has typically been performed by Kitware (Kitware, Inc), providing continuity and expertise); and (3) Kitware also offers commercial ITK support for a fee. Another example is that of Globus, which provides free support lists, operates a ticketing system [69], and guarantees subscribers a 1-business day response time on support tickets.

Surprisingly, free support is often available in a timely manner. One good example is the 3D Slicer, which had >13,000 forum posts in 2018, with an average response time of <2 days (or <8 hours during weekdays). For 3D Slicer, support may be provided either by the core developers or by experienced members of the user community. Public forums can be extremely active; for example, Bioconductor has >100 visitors per hour.

Ecosystem Collaboration

Ecosystem collaborations are usually organized by working groups, conferences, networks, and community forums. Limited public information is available on how well OSS projects collaborate with other projects.

Security

Security is important for biomedical software tools, as they are often used to manage and process patient data. To protect patient privacy, i2b2 provides secure remote access to patients in institutions through web services that anonymously list the number of patients in each institution [21]. Globus has maintained a strong security model for many years, using standards-based components and protocols that address message protection, authentication, delegation, and authorization for distributed infrastructures. Globus's authorization is based on well-established standards, such as OAuth 2 and OpenID Connect, and leverages a federated log-in system to allow user authentication using one of the many supported identity providers (eg, institutional identities, eRA Commons, ORCID, and Google [Google, Inc]). The Globus high assurance tier provides additional security controls to meet the higher authentication and authorization standards required for access to restricted data, such as protected health information. Data transfers can be encrypted using OpenSSL libraries, and communication channels with the Globus service are Transport Layer Security 1.2 encrypted.

Linux has strong security features and is widely used outside biomedical domains. The Linux kernel allows administrators to improve security at the lowest level by modifying the attributes of the kernel's operation, building additional security measures into the kernel to avoid common buffer overflow attacks, and setting different access restrictions for different kinds of users [70]. In addition, there are many Linux security extension enhancements, such as ExecShield and Position Independent Executable [71]. The other examined OSS use cases did not provide detailed information publicly about security. However, security enhancement should become the focus of future releases of research software.

Legal Concerns

Among the 10 OSS use cases, a popular licensing model is the GPL, which allows the distribution and sale of modified and unmodified versions but requires that all the copies be released under the same license and be accompanied by the complete corresponding source code. For example, Linux was released under GPL version 2, whereas R and tranSMART used GPL version 3.

It is also feasible to use different licensing models for different components of an OSS. For example, Bioconductor packages belong to multiple license groups: artistic license version 2, GPL, Massachusetts Institute of Technology (MIT), Berkeley Software Distribution (BSD), and creative commons licenses that have minimal requirements regarding how the software can be redistributed [72]. Globus also uses mixed licensing models. The client-side software is licensed under the Globus community license, which allows subscribers to access the source code for the purposes of code review and contribution, whereas the software operated by Globus as a service is not licensed.

Open-source licensing models used by the other OSS use cases include Apache 2 (OHDSI and ITK), Mozilla Public License version 2.0, with the Health care Disclaimer addendum (i2b2) [73], and GPL version 3 (tranSMART). REDCap requires a nonprofit end user license agreement between an institution and the Vanderbilt University, and its code base is not open to an individual developer. Finally, the 3D Slicer license, although generally highly permissive, is not a standard Open Source Initiative certified license. Instead, it is a custom license that was defined via coordination with the legal department of Brigham and Women's Hospital, which primarily aims to

mitigate liability risks because of the nature of the application (visualization and analysis in support of research applications on clinical images).

Financing

Of the 10 OSS use cases, 8 (80%) started with federal research funding. For example, Bioconductor began receiving the NIH National Human Genome Research Institute's support in 2003 and NCI/ITCR funding in 2014. The 3D Slicer has received direct or indirect support from many research grants (primarily NIH) over the course of several decades [74] but no sustained funding from any single source or program. Cytoscape received support from the National Institute of General Medical Sciences and the National Resource for Network Biology. REDCap received early support from the National Center for Research Resources. The early development of Globus was supported by the National Science Foundation and the Department of Energy, whereas more recent work on high assurance mechanisms has been supported by the NIH. Federal research funding is vital, as it encourages research on OSS to focus on scientific explorations and research ecosystem development. At the same time, although grants guarantee the researchers money to experiment, researchers still have to look for sustainable solutions beyond the grant cycle [29].

Industry and membership support are common in mature OSS cases. For example, premium Globus features (eg, data sharing, use reporting, and guaranteed support levels) are offered to institutions under an annual subscription, which is a flat annual fee based on the institutions' level of research activity. Linux continues to be supported by individual memberships (thousands of members) and annual corporate memberships (>1000 corporate members) [75]. The R Foundation is largely supported by members (membership fees from supporting persons, institutions, and benefactors) and *one-off donations*.

Multiple sponsor programs involving both academic and industry sponsors are also feasible. For example, ITK has continual funding from the NIH for maintenance to enable its free use and, at the same time, has commercial-grade support. OHDSI also has both private and public funding support. The i2b2 tranSMART Foundation has 4 sponsorship programs: contributing sponsors, corporate sponsors, sustaining sponsors, and event sponsors [76]. Through the tranSMART and the successor i2b2 tranSMART Foundation efforts, Keith Elliston and colleagues started Axiomedix (Axiomedix, Inc) in 2018 specifically to provide a commercial (for-profit) support mechanism for government-funded OSS. Axiomedix offers a 4-part business model that helps to support and sustain the open-source platforms: first, a commercial-grade software publishing and support model; second, a full-service solution offering for these supported platforms that includes installation, configuration, data loading, curation, and more; third, a software development and customization model (the Axiomedix Expert Network) that enables core open-source developers to take up contracts and consulting for customers; and finally, a model for developing new products and platforms that leverages open-source tools, a network of experienced open-source developers, and the knowledge of subject-matter experts to develop new open-source or commercial tools.

Marketing

The 10 OSS use cases have a variety of marketing channels, including the use of logos (3D Slicer, Globus, and i2b2 tranSMART), websites (3D Slicer, Bioconductor, Globus, and i2b2 tranSMART), mailing lists (Cytoscape, Globus, and i2b2 tranSMART), forums (3D Slicer, Cytoscape, and i2b2 tranSMART), Twitter (Twitter, Inc; 3D Slicer, Bioconductor, Cytoscape, Globus, and i2b2 tranSMART), LinkedIn (LinkedIn, Inc; Globus and i2b2 tranSMART), Facebook (Meta Platforms, Inc; i2b2 tranSMART), YouTube (Google, Inc; 3D Slicer, Bioconductor, and i2b2 tranSMART), Tumblr (Tumblr, Inc; Cytoscape), Vimeo (Vimeo, Inc; Cytoscape), and Pinterest accounts (Pinterest, Inc; Cytoscape).

Additional channels include conferences, workshops, and publications. For example, the ITK is introduced at medical imaging conferences. R gains market share through an *evangelist* approach among statisticians, data analysts, and others from the biomedical community. Moreover, surveys administered to collect user feedback also act as a form of marketing. For example, the 3D Slicer team conducts small-scale surveys on forums and collects feedback forms during training courses. Similarly, the Globus team conducts surveys during workshops and tutorials.

Dependency Hygiene

Of the 10 OSS (all except R), 9 (90%) have many dependencies on other packages. Bioconductor and OHDSI depend on many R packages, and REDCap depends on MySQL, whereas Cytoscape relies on external services, including cxMate. As dependencies may complicate installation and use, i2b2 provides Docker containers for easy installation [77]. Software models mainly provide dependency information through documentation, for example, installation guides; however, few models describe the license and security status of each dependency. Crichton [78] points out the potential danger of complicated dependencies, warning that "Blackbox can make it difficult to see that there are far fewer maintainers working behind the scenes at each of these open-source projects than what one might expect." Thus, it is critical to provide transparent information about the dependency tree of the code libraries. The 3D Slicer is a good example, as it provides an extensive list of dependencies that is publicly available.

Discussion

We discussed 10 representative OSS use cases that have demonstrated sustainable practices, particularly in the biomedical domain. Although not a comprehensive list, these examples highlight the following as essential attributes of successful OSS development: alignment with unmet scientific needs, a dedicated development team, a vibrant user community, a feasible licensing model, a sustainable financial model, and effective product management.

Alignment With Unmet Scientific Needs

At the inception of an OSS project, it must identify and meet important scientific needs instead of complying with mandatory rules or obtaining external financial rewards [79]. Meeting these needs gives the software its *soul*, that is, its unique identity. For

example, Cytoscape fulfills the need for a visualization tool to represent complex interactions among molecules, Bioconductor reduces the barrier to entry involved in the effective use and sharing of computational biology and bioinformatics tools [57], and Globus addresses the need for frictionless data transfer and sharing. As the scientific community's needs are diverse and dynamic, developers should consider the potential expansions beyond the first application and adopt a highly reusable infrastructure even at the initiation stage.

Dedicated Development Team

An OSS project should have a core development team, which has not only developed an initial version of the software but will also continue to be committed to future versions. The team is the *brain* of the software and its intellectual center. For example, Globus includes services for identity management, data transfer, data sharing, and group management; interfaces such as application programming interfaces, web apps, and a command-line client; and software to manage data access on >10 distinct storage platforms and file systems. Only a dedicated and highly experienced development team can put all these components together in a concerted fashion.

However, maintaining such teams can be difficult. According to Atiq et al [12], the motivations of developers usually include both intrinsic (eg, creativity and fun) and extrinsic aspects (eg, financial rewards, development of job-related skills, and peer recognition). Atiq et al [12] further pointed out that transparent and fair extrinsic rewards and effective and open communications among developers are key characteristics for ensuring the long-term sustainability of OSS projects.

More importantly, the whole research community needs to realize that the creation of a dedicated development team is incredibly difficult if that team cannot gain recognition for their contribution. Unfortunately, it is still true that, in academia, the effort invested in the development of software is often not recognized as important and is certainly viewed less favorably than traditional research activities.

Vibrant User Community

To be successful, an OSS project should also have a vibrant user community whose organizational structure and ongoing activities can facilitate communication both among and across the developer and user groups. This community would foster the *materialization* of the value of the software while specifying the functionality requirements for future versions. A vibrant user community represents the *heart* of the software, which drives the development cycle. For example, 3D Slicer and ITK have large and stable user bases, mainly in the radiology and biomedical imaging communities. OHDSI tools have large user bases in the clinical informatics and population health informatics communities. Moreover, we highly recommend engaging scientists outside the original team and involving a broad array of stakeholders. In addition, we support encouraging the *diaspora effect*, where postdoctorates and students who move on to other institutions continue using the software used or created by their original group.

It is also important to realize that the *users* of enterprise-level OSS are institutions, not individual researchers. In fact, Masys et al [79] defined successful adoption as at least 50% of the intended institutions adopting and implementing a tool. They suggested that, instead of a one-size-fits-all technical approach, developers should provide flexible local implementations and customizations, such as the optional use of terminology standards. This flexibility is essential for building a vibrant user community and facilitating successful adoption.

Feasible Licensing Model

A sustainable OSS project also needs a licensing model that fits the nature of the software, its distribution channel, and stakeholder interests. A licensing model resembles a *skeletal system*, providing a framework for the software to function legally.

OSS licensing generally falls into 4 categories: nonpermissive, weakly permissive, fully permissive, and noncompliant. Open-source licenses are evaluated as to whether they conform to the Open-Source Definition by the Open Source Initiative, a 501c3 nonprofit established to be a steward of open-source licenses.

Nonpermissive licenses, such as the GPL and the Affero GPL, not only allow commercial and noncommercial reuse but also require the release of all modified code and any external code linked to this code. The most well-known example is Linux, which is now under GPL version 2. Without the use of lawyers, its founder, Linus Torvalds, wrote a brief license stating that no fee may be charged for its distribution. As internet-based delivery systems were in an early stage of development, this move eliminated the *floppy drive mills* whereby individuals or companies could send copies of Linux to consumers for a fee. As the goal was not to allow others to make money on *free software distribution* at the time of writing, the model fit. When OSS code was to be modified or added on to, several open-source licenses were created and evolved. GPL version 3 (used by R and tranSMART) is the most restrictive open-source license, which requires that any enhancements (such as new features) incorporated into the software must be released along with the source code. Commercial software companies refer to the GPL version 3 as a *toxic license*. Once the software contains any GPL version 3 codes, its future licensing and that of all other software that carries it would be forever under the GPL version 3 license. The *infectiousness* keeps commercial companies away from using GPL code in their products; however, it could be one of the most important reasons why R is widely used and is successfully evolving. From our point of view, nonpermissive licenses may fit best for software that is fundamental to essential scientific discovery and highly used by researchers from very broad domains and where funding support may mostly come from noncommercial sources.

Weakly permissive licenses (eg, Mozilla Public License 2) allow commercial and noncommercial use and require release on a file-by-file basis for any modified code. Fully permissive licenses provide unrestricted reuse of code for commercial and noncommercial purposes. Fully permissive licenses include the Apache 2, MIT, and BSD licenses, among others. One of the main motivations of the popular Apache 2 license was to enable the ability to integrate open-source code into a project without having to release any enhancements to the code, that is, the

ability to *build on the shoulders of giants*. Finally, many projects that are considered open-source release codes under custom licenses are non–Open Source Initiative compliant. Thus, although these projects may make the code available, they cannot be considered open-source compliant.

There is a slow migration in the research software field toward fully permissive licenses because of limited commercial support. Elster [80] discusses how the license of research software may have an impact on obtaining industrial funding support. Many informatics technology companies choose research software with full permissive licenses over nonpermissive licenses, as nonpermissive licenses add restrictions to code reuse in commercial software, raising concerns about future commercialization. BSD license, as an example of a fully permissive license, allows the inclusion of open-source code in commercial code. On the other hand, some companies prefer nonpermissive licenses to fully permissive licenses, as they do not want their competitors to build commercial code on top of the OSS that those companies previously funded. Although this type of self-interested licensing prevailed in the early days of the software industry, the industry soon realized that having tens or hundreds of groups reinventing the same code was limiting the progress of the industry. As a result, there has been a wide and growing adoption of fully permissive licenses such as MIT and Apache [81].

Software licensing creates a binding agreement on the way a licensee may use or distribute the programs or codes. Just as a software-wrapped or click-through user licensing agreement is binding, so too is the use of OSS and code. When a research software is commercialized, a *free version* for academic use may be kept; however, if it is used outside the terms of that license, a *commercial license* must be purchased. Thus, the environment of the use of the software can play an important role in whether a user is in violation of the applicable license. A violation may result in harsh additional fees or even legal actions.

Sustainable Financial Model

An OSS project requires a sustainable financial model (formal or informal) that can keep the software and its user community moving forward. A sustainable financial model is a part of the *circulatory system*, supplying *blood* to sustain the software ecosystem. The i2b2 tranSMART, Globus, and Linux are excellent examples that leverage multiple types of sources to sustain software development.

The public–private partnership is becoming a feasible way to support an OSS project in the long term; however, the establishment of these partnerships may not be easy. Industry partners usually have concerns regarding profitable commercialization time. The public release of an OSS project, including its knowledge and source code, may allow the market competitors to catch up quickly, as opposed to traditional commercialized software business practices, where intellectual property is commonly concealed as long as possible. However, at the same time, an OSS project may quickly attract a large number of outside users and new developers whose contributions can improve the robustness of a product, enabling platform-based customizations across multiple institutions. Robust implementations and large user bases increase the commercial potential of OSS projects.

Along with the development of OSS, its financial model can change over time. Globus has tried a mix of several financial sustainability strategies: relying on grant-based federal funding, offering free OSS, forming an international research consortium, launching a commercial company, and forming an industry organization [82]. Globus found that many activities critical to sustaining software are outside the mission of federal funding agencies. Few developers have the freedom to contribute to a software project that addresses the research community's needs at large and does not directly advance the contributor's own mission. Contributors are not always able to provide ongoing maintenance or user support for the code they contributed, much less for the rest of the code base. As copies (*forks*) are maintained by separate teams, new features may no longer be shared with the entire community, and user requirements between the nonprofit research community and the industry do not always align. After 15 years, Globus pivoted to a sustainable model of providing free, cloud-based software-as-a-service to researchers and premium subscriptions for institutions. Focusing the primary software product on the needs of researchers and the revenue mechanism of creating value for resource providers is proving to be a viable financial model for sustaining Globus.

In the literature review section, we summarized a gamut of financial models for long-term software sustainability. Each approach has its own strengths and weaknesses. For example, community-based sustainability (eg, the not-for-profit organization model mentioned in the literature review), including appropriate forking of branch-development efforts, is in many ways ideal as it leverages the collective and continuous efforts of entire communities. However, it might not be appropriate for important niche areas of development; it might overemphasize broad adoption rather than quality, novelty, or significance, and it might not be able to leverage efforts that do not follow the same open-source licensing structure. Commercialization (eg, content licensing model, user subscription model, freemium model, razor and blades model), such as the adoption of software modules in clinical workstations, leverages a large pool of resources and software libraries in addition to creating a direct path to a broad user base willing to pay for it. However, commercialization is limited by proprietary restrictions and by its dependency on profit-making motives, which might not align well with biomedical significance or with *investment for the future* policies. Various infrastructure-based models (eg, macro research and development infrastructure model) can be effective ways to pool resources and avoid replication; however, they depend on a decision mechanism for the selection of the small percentage of software products that would be supported. Moreover, infrastructure-based models might be less prone to supporting innovation because of their not-so-dynamic nature. Various funding-based mechanisms (eg, national funding model and institutional support model) combine the advantages of dynamic selection and evolution of software products through the process of merit-based reviews. Unfortunately, they are limited by the harsh reality that existing funding is far less than the cost of

long-term maintenance of meritorious software, a situation that is unlikely to change in the foreseeable future.

Effective Product Management

Finally, an OSS project requires effective product management, which is a part of the *neural system*, enabling fast communications between the *brain* and other systems.

Roadmaps outline the development status of projects, including both the dates of past events and future events, so individuals can understand the speed, goals, and activities of specific projects, thereby improving sustainability through well-conceived deadlines and structures [83]. We found that 6 of the OSS cases on our list had well-designed roadmaps. The design of a road map is usually an evolving process that requires multiple rounds of internal discussions as well as extensive communication with the community of users and external developers. R and Linux offer a road map publicly. It is possible that these 2 OSS tools rely highly on the developer community's contributions, whereas the community makes its own decision about what it thinks is important, showing its partial resistance to central planning.

Although a road map designates the plans, it is the software release that shows the actual achievement. As OSS often involve the participation of a large number of external developers, the coordination of software releases can be more complicated. R provides a very good example by providing extensive support to facilitate external developers' package testing and release. Regarding the release strategy, OSS communities adopt either feature-based or time-based releases [84]. A feature-based release strategy is more often adopted by early-stage OSS projects. As an OSS project grows in size and complexity, it may move to time-based release, which helps prioritize development activities.

With regard to OSS quality assurance, a large user community may provide the project with good coverage in terms of bug hunting, performance, and scalability testing; however, most users do not consciously explore uncharted edge functionalities and thus leave certain bugs unfound [84]. Therefore, it is recommended to have professional testing and share a core bug report with the public through a public ticket tracker [85]. Moreover, an OSS project needs a version control system to coordinate release management, bug management, code stability and experimental development efforts, interdeveloper communication, and the authorization of changes by particular developers [86]. Public information shows that most of the 10 OSS use cases discussed go through rigorous testing.

As the *instruction manual for software* [87], documentation is essential in creating a sustainable community, as it allows users and external developers to rapidly become familiar with the software and use it for their own projects. Therefore, documentation is a key way of creating smoother internal transitions among generations of core developers. When familiarizing themselves with the OSS through the documentation is not enough for new users and external developers, specific support is essential to engage them, such as answering questions in a public forum. As mentioned in the *Results* section, all 10 OSS use cases provide comprehensive documentation and various types of support to users and new developers.

Strengths and Limitations

The selection of the examined software products was completed by 13 participants, comprising a group of people with rich knowledge of the sustainability of OSS tools and the promotion of industry partnerships. Although we conducted a comprehensive analysis of 10 aspects of the selected OSS use cases, there appears to be a risk of biasing the paper's findings toward the interests of the ITCR working group and overlooking potentially important sustainability models. Limited to publicly available information, we were not able to discuss failed OSS examples and important checkpoints. Our future goal is to conduct a survey of a much broader research community to continue these discussions.

In addition to the information discussed about general OSS aspects using the Nesbitt list [28], we would like to briefly discuss other important aspects of research software, such as scientific accuracy and reproducibility, compliance, and ethics and integrity. Rougier et al [88] defined reproducible software as the publishing of software and data as a product of the used software, its related data, and the articles involved. For software to be reproducible, its source code must be investigated, and its models must be documented thoroughly and precisely. Buck [89] explains that to improve reproducibility, transparency must be a top priority, despite the interference of high cost. To increase transparency, free OSS provides other scientists (besides software developers) with cheap options to validate their reported results and further apply this open science framework to other scientific research activities. Another aspect, compliance, is also critical for OSS, as the software may be incorporated into commercial uses, used to raise awareness about compliance, or used to display specific cases of noncompliance [90]. When distributed to external sources, the OSS licenses must be reviewed before compliance can be achieved (eg, for sharing, license fees, and compatibility purposes). Finally, ethics and integrity are essential for software in biomedical research. The use of OSS should allow researchers to meet the professional standards of practice, and the use of OSS must align with the 4 basic principles in the field: nonmaleficence, beneficence, autonomy, and justice.

Other Initiatives and Future Perspectives

In addition to the NCI ITCR, several informatics efforts across the NIH have also emphasized creating an approach to OSS sustainability. The National Center for Advancing Translational Sciences, Clinical and Translational Science Awards links with programs from the NIH Office of the Director, including the Big Data to Knowledge (BD2K) [91,92] and the Data Science program [93], and most recently, the All of US Precision Medicine Initiative [94]. BD2K is a trans-NIH initiative launched in 2013 to support the research and development of innovative and transformative approaches and tools that maximize and accelerate the integration of big data and data science into biomedical research. BD2K recognizes that software is a necessary part of any modern solution to biological problems. Representing the shared interest of the national Clinical and Translational Science Awards consortium, the

National Center for Data to Health is particularly interested in sustainability strategies for data management infrastructure, which again inevitably involves the sustainability of software tools revolving around clinical data.

Other countries, such as the United Kingdom and Germany, are also making national policies to improve software sustainability. Currently, the United Kingdom has developed a research and innovation road map and is using the *research and development system* as a connection to sources of funding that can flow to universities, research institutes, government laboratories, charities, and businesses [95]. The United Kingdom is moving toward minimizing bureaucracy in the public funding system to keep checks and approvals that will effectively manage public money and make informed decisions for the system. Moreover, the United Kingdom is increasing clarity and coherence in research and development funding to allow researchers to have confidence in long-term investments and enable agile funding to allow the system to tackle issues of national priority and urgency. For biomedicine, scholars in the United Kingdom recommend OSS in health care information systems to improve safety and effectiveness [96]. Similarly, Germany has created a more unified software policy [97] and has outlined the following recommendations: (1) in its foundation, research software must have an open-source code, as well as trustworthy, supportive, and appropriate infrastructure and infrastructure facilities; (2) senior researchers and research managers must develop good scientific practices, and there must be a general shift toward the acquisition of central licenses rather than commercial software and services; and (3) in the provision of research software, there must be a shift from the role of developer to the role of provider. There are still many challenges at the organizational and technical levels related to the development, use, and provision of research.

Looking forward, it will be important to learn from international governance examples and engage with other groups interested in sustainable software models. One notable community is the Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE) [98], a workshop series aimed at promoting sustainable research software by focusing on principles and best practices, careers, learning, and accreditation. The fourth WSSSPE created a group interested in writing white papers that focus on scientific environments and their implications, targeting developers and project managers of research software. Another notable community is the Science Gateways Community Institute, which provides consulting services for sustainability and business planning [99].

Conclusions

Overview

Our review of the existing sustainability models and 10 OSS use cases strongly confirms the importance of the 3 proposed future focus areas of the SIP-WG: (1) to develop a workflow or decision tree to support informed decision-making that is consistent with ITCR expectations and the future licensing needs of open-source tools; (2) to provide a consultancy service for the 10 sustainability aspects, especially governance, licensing, code quality, and community building, in collaboration with the ITCR program; and (3) to develop a collection of business model archetypes that can be used as starting templates to formally document the dissemination and sustainability plans for new software development initiatives. In addition, we stress on 5 important actions that should be considered in future ITCR activities, as described in the following sections.

Discussion of the Feasibility of Sustainability Models for ITCR Projects

An important agenda item of the SIP-WG's future work should be a discussion of the feasibility of various sustainability models for the many ITCR support projects, including nonprofit models (eg, the not-for-profit organization model, the national funding model, the infrastructure model, the institutional support model, and the donation model), and commercial models (eg, the distributed network model, the content licensing or industrial support model, the user subscription model, the freemium model, the split licensing model, the razor and blades model, the macro research and development infrastructure model, and the mixed model).

Exploration of the Potential Licensing Models

The licensing of research software will have a direct impact on public–private partnerships. A mixed licensing model may be the best way to strike a balance between free use (for broad use) and paid use (for funding support). Given the potential complexities of different OSS approaches, key stakeholders should consider the licensing structure of their software models as early as possible. Important decisions and changes must align well with the road map of software development and maintenance, as changing the licensing of existing projects can be very challenging. Once an open-source project integrates code from external contributors, it becomes logistically difficult to legally change the licensing on the code.

Provision of Reward Mechanisms to Enhance Stakeholders' Motivation to Focus on Sustainability

The WSSSPE community has pointed out the importance of enhancing stakeholders' motivation through credits and rewards [98]. Currently, the main credit given for developing a research OSS is through publications. Key contributors should be encouraged to list the creation of software resources on their resumes and further value the OSS in the grant funding review process. We should also provide reward mechanisms to fairly allocate credit to external developers who have contributed to successful expansions and adoptions. Finally, universities and research institutions should create viable career paths for researchers developing software in academia to encourage them to continuously work on research OSS development.

Establishment of a Central Library to Make OSS Visible and Reusable

In addition, we should consider establishing a central library to make ITCR-funded OSS more visible and reusable for a large number of biomedical researchers. The open-access library should index the OSS tools with brief descriptions of their functions and simple examples. This library should point to the latest version of each OSS tool. It would especially serve as a repository for retired OSS tools, which may have short-term difficulties in obtaining funding support. Ideally, this library

should be searchable, enabling something like a Google search for research OSS. When researchers have certain needs, they can first search within this library to find out if there is an existing tool available to meet their needs or if there is an existing tool they may expand upon to meet their needs.

Before establishing such a software library, we need to fully understand who the expected users of the library would be, what their incentive to use it would be, how often entries would be added and updated, whose responsibility would be to update the records, and what funding sources would support the future releases of a piece of software. Without continuous curation, there is an eventual risk that software libraries may become software graveyards.

Provision of Industry Standard Support

Finally, we should allocate consulting resources to research OSS projects (especially at the early stage of development), which can guide these projects to follow state-of-the-art industry standards on code quality control, ecosystem collaboration, security, and dependency hygiene.

Acknowledgments

The authors thank the ITCR program for supporting the SIP-WG authors who contributed to this manuscript. The authors especially want to thank Mervi Heiskanen for her support in the many meetings and the strategic coordination required to publish this manuscript. Finally, special thanks go to Juli Klemm, Program Director of the Center for Strategic Scientific Initiatives and Director of ITCR, for her strategic advice, inputs into the ITCR SIP-WG, and this important OSS sustainability model effort. The authors thank Dr Richard Boyce from the University of Pittsburgh for contributions to the OHDSI case study, Lee Liming and Vas Vasiliadis from the University of Chicago for contributions to the manuscript revisions, and Dr Rudi Pillich from the University of San Diego for providing guidance on the ITCR connectivity map. The authors thank the 2 reviewers for their insightful comments. The authors received the following funding support: MJB and YY, Centers for Disease Control and Prevention, National Institute for Occupational Safety & Health U24 OH009077; YY, National Library of Medicine K99LM013383; AF, NIH NCI U24 CA180918; MM, NIH U24CA180996; BDS, NCI-ITCR U01CA204826; GS, U24CA248010; SP, NIH-NIBIB grant P41 EB015902; and GQZ, ITCR R21 CA231904.

Authors' Contributions

MJB, YY, JRG, GS, GQZ, CD, and KOE defined the scope of the manuscript and drafted the initial sections. AF, IF, BER, and JCS provided critical reviews and contributed to the editing of the manuscript. AF, SP, JCFR, AL, GS, MM, JCS, MKD, IF, JRG, BER, DPT, JBZ, MJB, KOE YY, BDS, JVM, and GQZ reviewed OSS use cases and drafted Multimedia Appendix 1. SB contributed to the additional discussions about product management and the United Kingdom’s and Germany’s national policies of OSS. JC contributed to the additional discussions on licensing models.

Conflicts of Interest

KOE is a shareholder or investor in Ingentium, Inc; Axiomedix, Inc; Seneca Creek Research LLC; and Trazend, Inc; KOE holds advisory roles with Open-Source Pharma Foundation nongovernmental organization and i2b2 tranSMART Foundation and currently serves as the chief executive officer of PHEMI Systems Corp. MJB is a founder and has equity (stock) in SpIntellx, Inc.

Multimedia Appendix 1
Full descriptions of open-source software use cases.
[DOCX File , 207 KB - jmir_v23i12e20028_app1.docx]

References

1. About ITCR. National Cancer Institute Informatics Technology for Cancer Research. 2019. URL: https://itcr.cancer.gov/about-itcr [accessed 2021-11-09]
2. Funding opportunities. National Cancer Institute Informatics Technology for Cancer Research. 2019. URL: https://itcr.cancer.gov/funding-opportunities [accessed 2021-11-09]
3. Pratt D, Chen J, Pillich R, Rynkov V, Gary A, Demchak B, et al. NDEx 2.0: a clearinghouse for research on cancer pathways. Cancer Res 2017 Nov 01;77(21):e58-e61 [FREE Full text] [doi: 10.1158/0008-5472.CAN-17-0606] [Medline: 29092941]
4. Pillich R. ITCR Connectivity Map. URL: http://www.ndexbio.org/#/network/04c0a7e8-af92-11e7-94d3-0ac135e8bacf [accessed 2020-10-22]
5. Information Technology for Cancer Research (ITCR) initiative. NCI Hub. 2019. URL: https://ncihub.org/groups/itcr/sustainability_and_industry_partnership_wg [accessed 2021-11-09]
6. Kuchinke W, Ohmann C, Stenzhorn H, Anguista A, Sfakianakis S, Graf N, et al. Ensuring sustainability of software tools and services by cooperation with a research infrastructure. Per Med 2016 Jan;13(1):43-55 [FREE Full text] [doi: 10.2217/pme.15.43] [Medline: 29749867]

7. Chang V, Mills H, Newhouse S. Form open source to long term sustainability: review of business models and case studies. In: Proceedings of the UK e-Science All Hands Meeting 2007. 2007 Presented at: Proceedings of the UK e-Science All Hands Meeting 2007; Sep 10-13, 2007; Nottingham UK.
8. Aartsen W, Peeters P, Wagers S, Williams-Jones B. Getting digital assets from public-private partnership research projects through "the valley of death," and making them sustainable. Front Med (Lausanne) 2018;5:65 [FREE Full text] [doi: 10.3389/fmed.2018.00065] [Medline: 29594123]
9. Gabella C, Durinx C, Appel R. Funding knowledgebases: towards a sustainable funding model for the UniProt use case. F1000Res 2017;6:ELIXIR-2051 [FREE Full text] [doi: 10.12688/f1000research.12989.2] [Medline: 29333230]
10. Wikipedia. Business Models for Open-Source Software. 2019. URL: https://en.wikipedia.org/wiki/Business_models_for_open-source_software [accessed 2019-02-11]
11. Iaffaldano G, Steinmacher I, Calefato F, Gerosa M, Lanubile F. Why do developers take breaks from contributing to OSS projects?: a preliminary analysis. In: Proceedings of the 2nd International Workshop on Software Health. 2019 Presented at: 2nd International Workshop on Software Health; May 28, 2019; Montreal, Quebec, Canada. [doi: 10.1109/SoHeal.2019.00009]
12. Atiq A, Tripathi A. Impact of financial benefits on open source software sustainability. In: Proceedings of the Thirty Seventh International Conference on Information Systems. 2016 Presented at: Thirty Seventh International Conference on Information Systems; Dec 11-14, 2016; Dublin.
13. Jiménez RC, Kuzak M, Alhamdoosh M, Barker M, Batut B, Borg M, et al. Four simple recommendations to encourage best practices in research software. F1000Res 2017;6:ELIXIR-876 [FREE Full text] [doi: 10.12688/f1000research.11407.1] [Medline: 28751965]
14. Nyman L, Lindman J. Code forking, governance, and sustainability in open source software. Technol Innov Manag Rev 2013;3(1):7-12. [doi: 10.22215/timreview/644]
15. Nyman L, Mikkonen T, Lindman J, Fougère M. Perspectives on code forking and sustainability in open source software. In: IFIP International Conference on Open Source Systems. Berlin, Heidelberg: Springer; 2012:274-279.
16. Prins P, de Ligt J, Tarasov A, Jansen RC, Cuppen E, Bourne PE. Toward effective software solutions for big biology. Nat Biotechnol 2015 Jul;33(7):686-687. [doi: 10.1038/nbt.3240] [Medline: 26154002]
17. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J, Pujol S, et al. 3D Slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012 Nov;30(9):1323-1341 [FREE Full text] [doi: 10.1016/j.mri.2012.05.001] [Medline: 22770690]
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5(10):R80 [FREE Full text] [doi: 10.1186/gb-2004-5-10-r80] [Medline: 15461798]
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003 Nov;13(11):2498-2504 [FREE Full text] [doi: 10.1101/gr.1239303] [Medline: 14597658]
20. Chard K, Tuecke S, Foster I. Globus: recent enhancements and future plans. In: Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale. 2016 Presented at: XSEDE16 Conference on Diversity, Big Data, and Science at Scale; Jul 17 - 21, 2016; Miami, USA. [doi: 10.1145/2949550.2949554]
21. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010 Feb 26;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
22. Scheufele E, Aronzon D, Coopersmith R, McDuffie MT, Kapoor M, Uhrich CA, et al. tranSMART: an open source knowledge management and high content data analytics platform. AMIA Jt Summits Transl Sci Proc 2014 Apr 7;2014:96-101 [FREE Full text] [Medline: 25717408]
23. About ITK. ITK. URL: https://itk.org/ [accessed 2021-10-01]
24. About Linux. Linux. URL: https://www.linux.org/ [accessed 2021-11-09]
25. 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]
26. The R Project for Statistical Computing. 2021. URL: https://www.r-project.org/ [accessed 2021-10-01]
27. Obeid JS, McGraw CA, Minor BL, Conde JG, Pawluk R, Lin M, et al. Procurement of shared data instruments for Research Electronic Data Capture (REDCap). J Biomed Inform 2013 Apr;46(2):259-265 [FREE Full text] [doi: 10.1016/j.jbi.2012.10.006] [Medline: 23149159]
28. What Does A Sustainable Open Source Project Look Like? Libraries.io. 2017. URL: https://medium.com/libraries-io/what-does-a-sustainable-open-source-project-look-like-bf9b8cf824f8 [accessed 2021-11-09]
29. A handy guide to financial support for open source. GitHub. 2019. URL: https://github.com/nayafia/lemonade-stand [accessed 2021-10-01]
30. i2b2 tranSMART foundation amended and restated by-laws. URL: https://drive.google.com/file/d/0B8lizkKDeaKhTUF3QmNTTFk0ZnM/view [accessed 2017-04-25]

31. Amended and restated by-laws of i2b2 transmart foundation. i2b2 tranSMART. 2017. URL: http://roadmap-i2b2-transmart-pmc.hms.harvard.edu [accessed 2021-11-09]
32. Documentation/Labs/Slicer5-roadmap. 3D Slicer Roadmap. URL: https://www.slicer.org/wiki/Documentation/Labs/Slicer5-roadmap [accessed 2021-11-09]
33. Cytoscape product roadmap. Cytoscape Consortium. URL: https://cytoscape.org/roadmap.htm [accessed 2021-11-09]
34. Product roadmap. Globus Online. URL: https://www.globusworld.org/files/2010/02/120412-GW2012_Tuecke_Roadmap.pdf [accessed 2021-11-09]
35. ITK. ITK Roadmap 2005-2006. URL: https://itk.org/Wiki/ITK/Roadmap_2005_2006 [accessed 2020-09-01]
36. ITK Roadmap 2008-2009. ITK. URL: https://itk.org/Wiki/ITK_Roadmap_2008_2009 [accessed 2021-11-09]
37. ITK/Release 4/DICOM/Meeting Roadmap. ITK. URL: https://itk.org/Wiki/ITK/Release_4/DICOM/Meeting_2011.09.01_Roadmap [accessed 2021-11-09]
38. OHDSI architecture roadmap. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/web/wiki/doku.php?id=development:roadmap [accessed 2021-11-09]
39. CDM v6.0 Roadmap. Observational Health Data Sciences and Informatics. URL: https://forums.ohdsi.org/t/cdm-v6-0-roadmap/4809 [accessed 2021-11-09]
40. OHDSI WebAPI Roadmap. Observational Health Data Sciences and Informatics. URL: https://github.com/OHDSI/WebAPI/wiki/Roadmap [accessed 2021-11-09]
41. The Linux roadmap. Corbet. URL: https://lwn.net/Articles/114804/ [accessed 2021-11-09]
42. The ITK Software Guide. ITK. 2021. URL: https://itk.org/ItkSoftwareGuide.pdf [accessed 2021-11-09]
43. Linux guide. Wikibooks. URL: https://en.wikibooks.org/wiki/Linux_Guide [accessed 2021-11-09]
44. An Introduction to R. R-project. 2009. URL: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf [accessed 2021-10-26]
45. Welcome to 3D Slicer's Documentation. 3D Slicer. URL: https://slicer.readthedocs.io [accessed 2021-10-01]
46. 3D Slicer Wiki. GitHub. URL: https://github.com/Slicer/Slicer/wiki [accessed 2021-11-09]
47. Welcome to the i2b2 Community Wiki. i2b2 Community Wiki. URL: https://community.i2b2.org/wiki/pages/viewpage.action?pageId=342684 [accessed 2021-11-09]
48. Bioconductor Tutorial Part I. Bioconductor. 2003. URL: https://bioconductor.org/help/course-materials/2003/MGED6/MGED6I.pdf [accessed 2021-11-09]
49. Harness the power of the Globus research data management cloud. Globus Docs. URL: https://docs.globus.org/ [accessed 2021-11-09]
50. Cytoscape tutorials. GitHub. URL: https://github.com/cytoscape/cytoscape-tutorials/wiki [accessed 2021-11-09]
51. tranSMART Platform Tutorials. i2b2 tranSMART Foundation. URL: https://transmartfoundation.org/transmart-platform-tutorials/ [accessed 2021-11-09]
52. 3D Slicer Nightly Tutorials. URL: https://www.slicer.org/wiki/Documentation/Nightly/Training [accessed 2020-03-12]
53. OHDSI. 2019 Tutorials – OMOP Common Data Model and Standardized Vocabularies. Observational Health Data Sciences and Informatics. 2019. URL: https://www.ohdsi.org/2019-tutorials-omop-common-data-model-and-standardized-vocabularies/ [accessed 2021-11-09]
54. REDCap Tutorial. YouTube. URL: https://www.youtube.com/playlist?list=PLrnf34ZtZ9FohXnuD6KmwVb1xgbGDu3rY [accessed 2021-11-09]
55. 3D Slicer YouTube Tutorials. YouTube. URL: https://www.youtube.com/channel/UC8vxI0-dEWrw0_tBF-v8xGA [accessed 2021-11-09]
56. Cytoscape YouTube Channel. YouTube. URL: https://www.youtube.com/channel/UCv6auk9FK4NgXiXiqrDLccw [accessed 2021-11-09]
57. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 2015 Feb;12(2):115-121 [FREE Full text] [doi: 10.1038/nmeth.3252] [Medline: 25633503]
58. Cytoscape app ladder. GitHub. 2019. URL: https://github.com/cytoscape/cytoscape/wiki/Cytoscape-App-Ladder [accessed 2021-11-09]
59. The R Manuals. CRAN. URL: https://cran.r-project.org/manuals.html [accessed 2021-11-09]
60. 3D Slicer Preview. CDash. URL: http://slicer.cdash.org/index.php?project=SlicerPreview [accessed 2021-11-09]
61. Releasing a package. R Packages. URL: http://r-pkgs.had.co.nz/release.html [accessed 2021-11-09]
62. Documentation/Nightly/Developers/Style Guide. 3D Slicer. URL: https://www.slicer.org/wiki/Documentation/Nightly/Developers/Style_Guide [accessed 2021-11-09]
63. R Internals for R Version 3.6.3. 2020. URL: https://cran.r-project.org/doc/manuals/R-ints.html [accessed 2020-03-12]
64. OHDSI. OHDSI Forums. Observational Health Data Sciences and Informatics. URL: https://forums.ohdsi.org/ [accessed 2021-11-09]
65. OHDSI Github Repository. Observational Health Data Sciences and Informatics. URL: https://github.com/ohdsi/ [accessed 2021-11-09]
66. Linux Help Center. Linux Foundation. URL: https://support.linuxfoundation.org/ [accessed 2021-11-09]

67. Cytoscape Helpdesk homepage. Cytoscape Helpdesk. URL: https://groups.google.com/forum/#!forum/cytoscape-helpdesk [accessed 2021-11-09]
68. ITK discussion forum. ITK. URL: https://discourse.itk.org [accessed 2021-11-09]
69. Connect with Globus on social media. Globus. URL: https://www.globus.org/contact-us [accessed 2021-11-09]
70. Mookhey K, Burghate N. Linux: Security, Audit and Control Features. East Golf Road, Schaumburg, Illinois: ISACA; 2005.
71. UNIX/LINUX Security Features. Winsock. URL: https://www.tenouk.com/linuxunixsecurityfeatures.html [accessed 2021-11-09]
72. How R Packages on bioconductor are licensed. CRAN. URL: http://seankross.com/2016/08/02/How-R-Packages-are-Licensed.html [accessed 2021-11-09]
73. The i2b2 MPL 2.0 License with Healthcare Disclaimer Addendum. i2b2 Web Client. URL: https://community.i2b2.org/wiki/display/webclient/The+i2b2+MPL+2.0+License+with+Healthcare+Disclaimer+Addendum [accessed 2021-11-09]
74. Documentation/4.x/Acknowledgments. 3D Slicer. URL: https://www.slicer.org/wiki/Documentation/4.x/Acknowledgments [accessed 2021-11-09]
75. Linux Foundation Membership. The Linux Foundation. URL: https://www.linuxfoundation.org/membership/ [accessed 2021-11-09]
76. i2b2 tranSMART Foundation Sponsorship Programs. URL: https://transmartfoundation.org/contributing-sponsors-program/ [accessed 2021-11-09]
77. Wagholikar KB, Dessai P, Sanz J, Mendis ME, Bell DS, Murphy SN. Implementation of informatics for integrating biology and the bedside (i2b2) platform as Docker containers. BMC Med Inform Decis Mak 2018 Jul 16;18(1):66 [FREE Full text] [doi: 10.1186/s12911-018-0646-2] [Medline: 30012140]
78. Open source sustainability. Tech Crunch. 2018. URL: https://techcrunch.com/2018/06/23/open-source-sustainability/ [accessed 2021-11-09]
79. Masys DR, Harris PA, Fearn PA, Kohane IS. Designing a public square for research computing. Sci Transl Med 2012 Aug 29;4(149):149fs32 [FREE Full text] [doi: 10.1126/scitranslmed.3004032] [Medline: 22932220]
80. Software for science: some personal reflections. arXiv. 2013. URL: https://arxiv.org/abs/1309.2357 [accessed 2021-11-09]
81. Open source licenses: trends and predictions. White Source. 2021. URL: https://resources.whitesourcesoftware.com/blog-whitesource/open-source-licenses-trends-and-predictions [accessed 2021-11-09]
82. Guest Blog: Globus and the Path to Sustainability. Science Gateways. 2019. URL: https://sciencegateways.org/-/guest-blog-globus-and-the-path-to-sustainability?inheritRedirect=true&redirect=/community/blog [accessed 2021-11-09]
83. Fogel K. Chapter 2: getting started. In: Producing Open Source Software: How to Run a Successful Free Software Project. Second edition. Under the CreativeCommons Attribution-ShareAlike (4.0) license: Copyright © 2005-2021 Karl Fogel; 2020.
84. Teixeira J. Release early, release often and release on time. An empirical case study of release management. In: Proceedings of the IFIP International Conference on Open Source Systems. 2017 Presented at: IFIP International Conference on Open Source Systems; May 22-23, 2017; Buenos Aires, Argentina. [doi: 10.1007/978-3-319-57735-7_16]
85. Fogel K. Chapter 5: participating as a business, non-profit, or government agency. In: Producing Open Source Software: How to Run a Successful Free Software Project. Second edition. Under the CreativeCommons Attribution-ShareAlike (4.0) license: Copyright © 2005-2021 Karl Fogel; 2020.
86. Fogel K. Chapter 3: techinical infrastructure. In: Producing Open Source Software: How to Run a Successful Free Software Project. Second edition. Under the CreativeCommons Attribution-ShareAlike (4.0) license: Copyright © 2005-2021 Karl Fogel; 2020.
87. Eghbal N. Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure. New York, NY: Ford Foundation; 2016.
88. Rougier N, Hinsen K, Alexandre F, Arildsen T, Barba LA, Benureau FC, et al. Sustainable computational science: the ReScience initiative. PeerJ Comput Sci 2017;3:e142 [FREE Full text] [doi: 10.7717/peerj-cs.142] [Medline: 34722870]
89. Buck S. Solving reproducibility. Science 2015 Jun 25;348(6242):1403. [doi: 10.1126/science.aac8041]
90. Schoettle H. Open source license compliance-why and how? Computer 2019 Aug;52(8):63-67. [doi: 10.1109/mc.2019.2915690]
91. Big Data to Knowledge. National Institutes of Health Office of Strategic Coordination The Common Fund. URL: https://commonfund.nih.gov/bd2k [accessed 2021-11-09]
92. Bourne PE, Bonazzi V, Dunn M, Green ED, Guyer M, Komatsoulis G, et al. The NIH Big Data to Knowledge (BD2K) initiative. J Am Med Inform Assoc 2015 Nov;22(6):1114 [FREE Full text] [doi: 10.1093/jamia/ocv136] [Medline: 26555016]
93. Data science at NIH. National Institutes of Health Office of Data Science Strategy. URL: https://datascience.nih.gov/ [accessed 2021-11-09]
94. The precision medicine initiative cohort program – building a research foundation for 21st century medicine. NIH. 2015. URL: https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf [accessed 2021-11-09]

95. UK research and development roadmap. Gov.UK. 2020. URL: https://www.gov.uk/government/publications/uk-research-and-development-roadmap [accessed 2021-11-09]
96. Reynolds CJ, Wyatt JC. Open source, open standards, and health care information systems. J Med Internet Res 2011 Feb 17;13(1):e24 [FREE Full text] [doi: 10.2196/jmir.1521] [Medline: 21447469]
97. Recommendations on the Development, Use and Provision of Research Software. Zenodo. 2018. URL: https://zenodo.org/record/1172988#.X9Q1UdhKiF [accessed 2021-11-09]
98. Katz D, Choi S, Lapp H, Maheshwari K, Löffler F, Turk M. Summary of the first Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE1). J Open Res Softw 2014;2(1):e6. [doi: 10.5334/jors.an]
99. Science Gateways Community Institute. URL: https://sciencegateways.org/gateway-support/consulting-services [accessed 2021-11-09]

Abbreviations

BD2K: Big Data to Knowledge
BSD: Berkeley Software Distribution
GPL: general public license
i2b2: Informatics for Integrating Biology and the Bedside
ITCR: Informatics Technology for Cancer Research
ITK: Insight Toolkit
MIT: Massachusetts Institute of Technology
NCI: National Cancer Institute
NIH: National Institutes of Health
OHDSI: Observational Health Data Sciences and Informatics
OSS: open-source software
REDCap: Research Electronic Data Capture
SIP-WG: sustainability and industry partnership working group
WSSSPE: Workshop on Sustainable Software for Science: Practice and Experiences

Edited by R Kukafka; submitted 17.05.20; peer-reviewed by N Chue Hong, S Hettrick; comments to author 23.07.20; revised version received 14.12.20; accepted 23.09.21; published 02.12.21.

Please cite as:
Ye Y, Barapatre S, Davis MK, Elliston KO, Davatzikos C, Fedorov A, Fillion-Robin JC, Foster I, Gilbertson JR, Lasso A, Miller JV, Morgan M, Pieper S, Raumann BE, Sarachan BD, Savova G, Silverstein JC, Taylor DP, Zelnis JB, Zhang GQ, Cuticchia J, Becich MJ
Open-source Software Sustainability Models: Initial White Paper From the Informatics Technology for Cancer Research Sustainability and Industry Partnership Working Group
J Med Internet Res 2021;23(12):e20028
URL: https://www.jmir.org/2021/12/e20028
doi:10.2196/20028
PMID:34860667

Viewpoint

Analyzing Patient Trajectories With Artificial Intelligence

Ahmed Allam[1,2], PhD; Stefan Feuerriegel[3,4,5], PhD; Michael Rebhan[3], PhD; Michael Krauthammer[1,2,6], MD, PhD

[1]Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

[2]Biomedical Informatics, University Hospital of Zurich, Zurich, Switzerland

[3]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[4]ETH Artificial Intelligence Center, ETH Zurich, Zurich, Switzerland

[5]Ludwig Maximilian University of Munich, Munich, Germany

[6]Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, United States

Corresponding Author:
Stefan Feuerriegel, PhD
Ludwig Maximilian University of Munich
Geschwister-Scholl-Platz 1
Munich, 80539
Germany
Phone: 49 8921806790
Email: feuerriegel@lmu.de

Abstract

In digital medicine, patient data typically record health events over time (eg, through electronic health records, wearables, or other sensing technologies) and thus form unique patient trajectories. Patient trajectories are highly predictive of the future course of diseases and therefore facilitate effective care. However, digital medicine often uses only limited patient data, consisting of health events from only a single or small number of time points while ignoring additional information encoded in patient trajectories. To analyze such rich longitudinal data, new artificial intelligence (AI) solutions are needed. In this paper, we provide an overview of the recent efforts to develop trajectory-aware AI solutions and provide suggestions for future directions. Specifically, we examine the implications for developing disease models from patient trajectories along the typical workflow in AI: problem definition, data processing, modeling, evaluation, and interpretation. We conclude with a discussion of how such AI solutions will allow the field to build robust models for personalized risk scoring, subtyping, and disease pathway discovery.

(J Med Internet Res 2021;23(12):e29812) doi:10.2196/29812

KEYWORDS

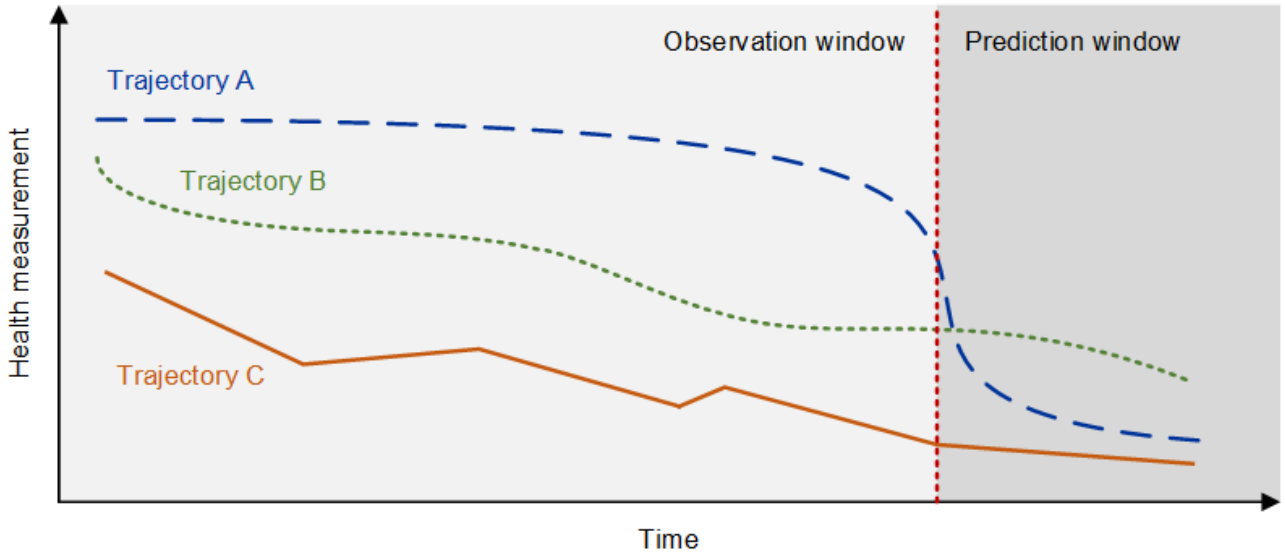
patient trajectories; longitudinal data; digital medicine; artificial intelligence; machine learning

Introduction

Digital medicine facilitates broad access to large volumes of patient data, typically through recordings of health events over time. For example, electronic health records store the history of a patient's diagnoses, medications, laboratory values, and treatment plans [1-3]. Wearables collect granular sensor measurements of various neurophysiological body functions over time [4-6]. Intensive care units (ICUs) monitor disease progression via continuous physiological measurements (eg, electrocardiograms) [7-10]. As a result, patient data in digital medicine are regularly of longitudinal form (ie, consisting of health events from multiple time points) and thus form *patient trajectories*.

Analyzing patient trajectories provides opportunities for more effective care in digital medicine [2,7,11]. Patient trajectories encode rich information on the history of health states that are also predictive of the future course of a disease (eg, individualized differences in disease progression or responsiveness to medications) [9,10,12]. As such, it is possible to construct patient trajectories that capture the entire disease course and characterize the many possible disease progression patterns, such as recurrent, stable, or rapidly deteriorating disease states (Figure 1). Hence, modeling the patient trajectories allows one to build robust models of diseases that capture disease dynamics seen in patient trajectories. Here, we replace disease models with data from only a single or a small number of time points by disease models that account for the longitudinal nature of patient trajectories, thus offering vast potential for digital medicine.

Figure 1. Analyzing patient trajectories with artificial intelligence in digital medicine.



Several studies have previously introduced artificial intelligence (AI) in medicine for practitioners [13,14]. Some studies review potential medical applications that could benefit from AI [15,16], whereas others review specific methods (eg, deep learning [17,18]) or specific data types (eg, electronic health records [18] and medical images [19]). Some studies suggest reporting guidelines [20,21] or discuss the integration of AI into medical practice [22]. Our research contributes to the literature by discussing AI-powered digital medicine based on patient trajectories.

Patient trajectories refer to time-resolved representations of patient health events across multiple time points (eg, hospitalization or treatment events, sensor measurements from wearables, and physiological measurements from ICUs). AI-powered analysis of patient trajectories allows for an assessment of the heterogeneity of patient disease courses. In Figure 1, trajectory A predicts a sudden but sharp decline in a health state, whereas trajectories B and C depict 2 types of progressively declining disease states. Analyzing trajectory-like representations of patient data can generate new insights for better care in digital medicine.

To unlock the value of patient trajectories for digital medicine, there is a need for new AI solutions that can deal with time-resolved sequential data consisting of multiple health events. Although many models from the area of AI have become standard in digital medicine (eg, deep learning [18]), a naïve application of such models might not be effective when modeling the longitudinal nature of patient trajectories. Instead, this requires customized approaches. For example, in a study by Alaa et al [23], a direct application of deep learning has been found to be outperformed in terms of both predictive accuracy and interpretability when one instead uses a carefully engineered sequential model (ie, referred to as Hawkes process) that treats the time between medical events as informative for the course of the disease. On the basis of this background, we discuss challenges and solutions for AI that are unique to analyzing patient trajectories. Specifically, we examine the implications for developing disease models from the patient trajectories along the typical workflow in AI: (1) problem definition, (2) data processing, (3) modeling, (4) evaluation, and (5) interpretation, as detailed in the following section.

Applying AI to Patient Trajectories

Problem Definition

Applying AI to patient trajectories is relevant for different objectives in digital medicine (Table 1). One objective is *risk scoring* [3,23,24], where patient trajectories are leveraged to predict patient outcomes. Here, the rationale is that the predictions based on health measurements from patient trajectories with multiple time points have greater predictive power than the predictions from those with a single or a few time points. For instance, risk scoring in ICUs becomes more accurate when traditional scores (eg, Acute Physiology and Chronic Health Evaluation II and Simplified Acute Physiology Score) are replaced with AI-based predictions incorporating data from patient trajectories [9,10,25]. Similarly, for cardiovascular diseases, existing risk scores (eg, the Framingham risk score that predicts the 10-year risk of developing coronary heart disease) become more accurate when replaced by AI solutions that work with longitudinal patient data [12]. These examples show that the underlying patient trajectory provides rich, granular insights into the disease dynamics that can then be captured by AI solutions for trajectory analysis. Therefore, additional patient information, such as past medications, comorbidities, or other risk factors, can be considered. For instance, in the context of cardiovascular diseases, it might be informative for risk scoring to analyze the patient's past journey, which comprises whether the patients have been prescribed nicotine replacements and *when* (eg, only recently or several years ago). Different patient outcomes including mortality, hospital readmission, hospital length of stay, disease onset, disease severity, or adverse drug reactions can be of interest in risk scoring. The risk score can then inform treatment planning (or in general, assess the patients' needs). In addition, AI solutions can further generate insights for defining (early) disease states.

Table 1. Overview of different objectives in artificial intelligence–based trajectory analysis.

Objective	Description	Examples	Selected references
Risk scoring	The objective is to estimate the likelihood of future health outcomes (eg, mortality, readmission, and adverse drug reactions)	• Predict the 10-year risk of developing coronary heart disease for patients as in the Framingham risk score • Predict the need for an intensive care unit in an emergency ward through measurements from wearables	[3,17,23,26-32]
Subtyping	The objective is to cluster the patient cohort into different disease dynamics (ie, subtyping) while accounting for the longitudinal form of patient trajectories	• Cluster disease progressions into "recurrent course" and "progressive decline"	[26]
Pathway discovery	The objective is to detect clinically meaningful subpatterns in patient trajectories	• Identify frequent patterns in patient trajectories that are indicative of disease onset	[1,33,34]

Here, we see several paths for risk scoring based on patient trajectories. First, to ensure high accuracy, AI-based risk scores must be tailored to each patient outcome while considering the desired forecast horizon and the patient cohort. So far, there are several studies that showcase the successful use of AI-based risk scores [35,36]; however, there is a need to develop other risk scores, especially for the settings in which AI-based risk scores are scarce or not yet available (eg, predicting the transition from prediabetes to diabetes or predicting specific adverse reactions to medication). Second, the AI-based risk scores will need to be integrated more extensively in clinical practice. Third, the risk scores should be extensively combined with approaches for explainability or interpretability, which allow the derivation of clinically relevant insights from patient trajectory data (eg, which information in a patient trajectory is a risk factor). Finally, if one includes data on treatments in the risk scoring model, one may infer the expected individualized treatment effect and eventually guide the treatment selection [37-42]. Here, we see further potential to transition from a purely predictive approach (ie, what is the expected risk level) to a prescriptive approach (ie, what treatment do we expect to reach a desired patient outcome). However, many AI solutions for estimating individualized treatment effects from patient trajectories have recently emerged [37-42] without being tailored to specific disease settings and patient cohorts. Therefore, further research at the interface to digital medicine should be a priority that will eventually yield effective and robust implementations for clinical practice.

Another objective of AI in digital medicine is *subtyping*, where AI can understand the heterogeneity observed in patient trajectories and identify the corresponding digital markers. A simple approach is to cluster the different patient trajectories (ie, subtyping) to match patients with similar disease dynamics, clinical pathways, or care patterns [26]. As a practical benefit, subtyping can support clinical tasks related to cohort building and, for instance, can provide patient stratification (eg, to define a cluster of patient trajectories that serves as an inclusion criterion for a clinical trial). However, subtyping requires a suitable notion of patient similarity, which can be challenging to define mathematically because of the longitudinal form of patient trajectories. Thus, it is crucial not only to cluster risk factors at baseline but also to find mathematical approaches that account for the temporal nature of the trajectories (ie, time-series clustering). This allows clinical practice to identify subgroups or subtypes based on the underlying disease dynamics (eg, to distinguish subgroups with a recurrent course vs a progressive decline). We expect an added value from comparing different subtyping approaches in terms of their relative strengths (eg, generated insights) for future research. On this basis, digital medicine could develop a principled procedure for defining *patient trajectory similarity* in the context of subtyping.

A related objective is *pathway discovery***,** where patterns in patient trajectories should be detected [1,33]. For instance, 1 application analyzes time series with laboratory measurements from patients with hepatitis B and C to discover frequent patterns indicative of liver damage [34]. This application can help to understand the underlying course of diseases and identify short- and long-term patterns (ie, motifs) in patient trajectories.

Depending on the objective and explicit assumptions, implications arise regarding the AI workflow and thus highlight the importance of selecting an appropriate modeling strategy. Additional details are provided in the following sections.

Data Processing

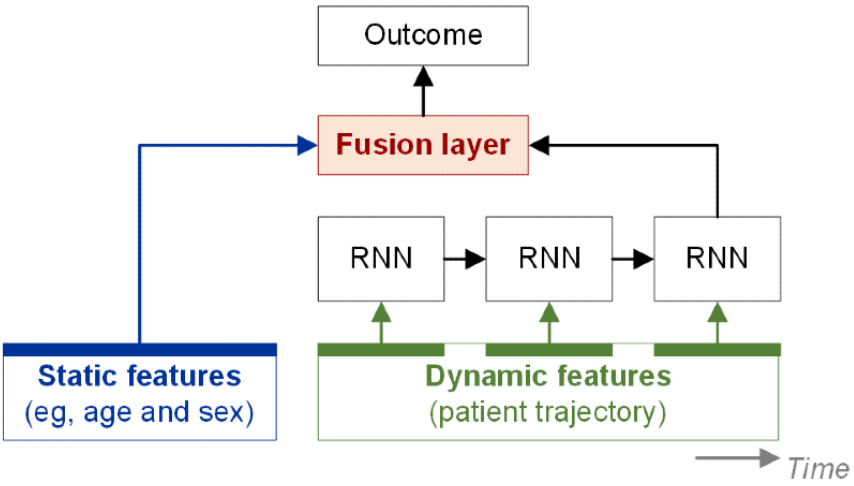
A fundamental question is concerned with data collection as this defines how time is encoded in the data. The example of nicotine replacement suggests that we should know whether such a medical event was recent or several years ago. This illustrates where we are now: when we have longitudinal data, it is often a matter of whether a medical event happened *recently* or *a while back*. Thus, the underlying time (or the underlying time lag) must be carefully considered to capture the longitudinal dimension of patient trajectories correctly. This is currently a challenge when considering the survey designs (eg, for risk scores). For example, one survey may ask a patient whether an event occurred last month or earlier, whereas another survey may ask to consider events that occurred within the last 12 months or earlier. As such, the meaning of *recent* may be inconsistent across survey designs. As a way forward, it will be necessary to develop a more consistent understanding, ideally involving data collection that considers precise time stamps (eg, by leveraging electronic health records).

The AI-based trajectory analysis often combines data from patient trajectories and baseline variables that describe risk

factors at the patient level (eg, sociodemographic, genomic data, or multimodal data). To combine sequential and static baseline data, tailored AI solutions will need to be developed [27]. Figure 2 shows an example of this approach. The literature shows larger variability regarding the modeling approaches; hence, further evaluations are needed to inform an effective approach.

Figure 2 shows an AI-based trajectory analysis in which a fusion layer combines the static (eg, age or sex) and dynamic features (eg, health measurements over time). The dynamic features have a longitudinal form and are processed by a sequential model (here, a recurrent neural network [RNN]).

Figure 2. Example of artificial intelligence–based trajectory analysis. RNN: recurrent neural network.



Data from patient trajectories are often complex and high-dimensional, which presents difficulties to humans and AI solutions for making accurate inferences. As a remedy, one can apply approaches that map patient trajectories onto a lower-dimensional representation that is eventually more meaningful. A simple analogy from medical practice is when one simplifies the age in years of a patient into a binary yes or no flag indicating whether a patient is above a critical age threshold. Here, we see a particular value for representation learning (eg, embeddings [31]) tailored to the unique characteristics of longitudinal health time series from clinical practice. Such AI solutions must be effective in dealing with the high-dimensional nature of medical data (clinical, genetic, social data, etc), avoid overfitting, and overcome the *curse of dimensionality* in the analysis. Mathematically, the idea of learning lower-dimensional representations is linked to manifold learning, for which embeddings are a special case [43].

Another challenge that further limits the use of AI-based trajectory analysis in medicine is data access and sharing. The current system is characterized by having data locked in *silos* where each hospital or health care institution limits access to their data and requires *lots of bureaucratic* work from researchers before allowing them to study and analyze the data. However, many initiatives are circumventing this status quo, such as the Observational Health Data Sciences and Informatics program, an international network of researchers aiming to provide reusable, collaborative, and reliable open-source solutions for large-scale health analytics [44]. Notably, there is interest in compiling extensive observational studies combining the electronic health record data from diverse health care organizations using standards to (1) design meaningful randomized controlled trials, (2) test clinical hypotheses on observational data, and (3) gain a better understanding of population characteristics, facilitated through framework efforts such as the Observational Health Data Sciences and Informatics program [44].

Furthermore, in the last couple of years, there has been an increasing number of studies focused on federated learning [45-47] that allows for AI algorithms to operate on decentralized data sets/systems in a privacy-preserving manner. In federated learning, the underlying algorithms (eg, FedAvg [48] and FedProx [49]) use the data stored in silos at different local sites for iteratively training a *global/central* model from a set of *local* models trained separately at each local site to perform prediction and classification tasks [48,50]. At the interface to health care, more research is being conducted that uses federated learning [45,51] for tasks such as clinical note phenotyping (ie, clinical natural language processing [52]) or predicting patient mortality in ICUs [53,54]. Here, we particularly point to the recent attempts to develop such approaches for patient trajectories. Examples include clustering patients through community-based federated machine learning for in-hospital mortality and length of stay prediction [55] or privacy-preserving patient similarity learning [56]. Federated learning may be further supported by secure hardware implementations, often with little computational overhead (eg, referred to as trusted execution environments [47]).

Moreover, there is more research on the privacy-preserving aspect of the technology, such as the differential privacy framework [57] (ie, applied to the model parameters), homomorphic encryption [58], and data anonymization techniques offering a *defensive level of privacy* as required by the General Data Protection Regulation and the Health Insurance Portability and Accountability Act [59]. Although these provide valuable tools for developers, we foresee more research that tailors them to the context of patient trajectories (eg, by offering sequential models for longitudinal data). More importantly, the availability of software packages [60,61] that allow both simulation of federated learning scenarios and their deployment

in real clinical settings will accelerate the adoption of federated or distributed learning approaches and open a wide array of research exploration and experimentation possibilities. This will eventually yield longitudinal trajectory analyses that span patient journeys across multiple hospitals or health care institutions.

Modeling

Fundamentally, analyzing patient trajectories requires AI solutions that can effectively handle sequential data structures that can vary in length (ie, from a few time points to multiple seconds, minutes, days, months, and year time windows). Hence, AI-based trajectory analysis must carefully adapt to the sequential structures by choosing appropriate modeling approaches.

In *risk scoring*, predictions from patient trajectories are often based on neural networks that are tailored to sequential data structures. These include tailored RNNs [3,17,27-29] owing to their strength in modeling long-term dependencies. One example of RNNs is the long short-term memory networks that iteratively process a time series with physiological measurements and aim to learn a lower-dimensional representation of the complete time series, regardless of its length, in their internal code layer. We can then use this lower-dimensional representation to predict patient outcomes from the patient's trajectory. Gated recurrent units proceed analogously but have a more parsimonious structure that, in medical applications, may help in obtaining robust predictions (eg, with a lower risk of overfitting for small-sized data sets) [8,27,62]. Recently, digital medicine has also processed the patient trajectories through transformer networks [30]. Transformer networks involve attention layers that learn to weigh different parts in a patient trajectory differently while optimizing for the outcome prediction and may therefore outperform other RNNs. In our view, a particular benefit of transformer networks is that the developers from digital medicine can train them in using semisupervised learning. One can use a set of patient trajectories *without* observing the patient outcomes to learn an abstract representation (ie, via unsupervised pretraining). Subsequently, one can customize the transformer network to predict a specific patient outcome (ie, via supervised fine-tuning). Semisupervised learning often facilitates more efficient learning when the number of available observations with patient outcomes is comparatively low.

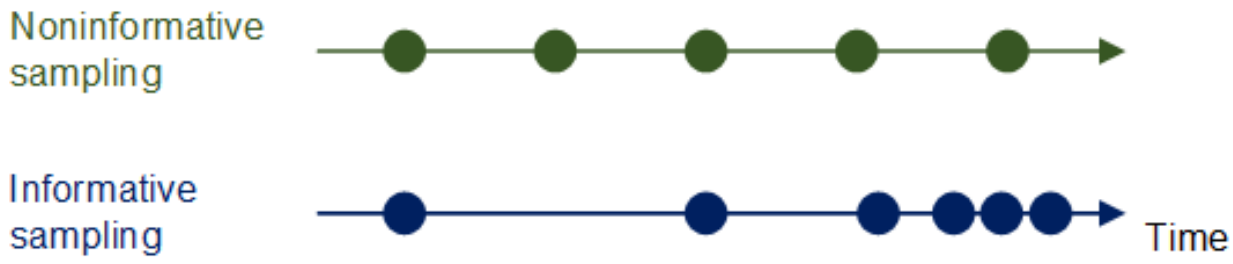
In risk scoring, other common prediction approaches are probabilistic models (eg, Markov models, point processes, and Gaussian processes) [23,32]. Here, we see several benefits for patient trajectory analyses in clinical settings. Probabilistic models often have a more parsimonious structure than the out-of-the-box neural networks, which facilitates efficient learning and reduces the risk of overfitting in data-scarce settings. In addition, such a parsimonious structure can facilitate interpretation by clinical practitioners. Probabilistic models can be naturally extended by latent structures (eg, hidden Markov models [63-67]), where latent states capture different trajectory phases and further improve interpretability. For instance, in the context of alcoholism treatment, patient trajectories have been modeled to undergo phases of *abstinence*, *moderate drinking*, and *heavy drinking*, each of which is captured by a separate latent state. In our view, such a latent structure represents a natural way to describe the different patterns in patient trajectories (eg, acute vs stable phases) and, more importantly, relates model characteristics to established clinical terminology. In the future, we expect hybrid models that combine the benefits of probabilistic modeling and neural learning (eg, deep Markov models [25]). The former benefits from theory-informed, interpretable structures that account for different disease states, whereas the latter are particularly effective for long-term dependencies.

Across these risk models, it is further essential to consider the timing of the health measurement. Trajectories may consist of health recordings in equally spaced time intervals (eg, uniformly sampled every minute in ICUs or yearly intervals in patient registries). Often, they contain irregular time intervals, reflecting nonuniform and patient-specific interactions with the health care system. As a result, the sampling might be informative of the disease state (Figure 3). For instance, health professionals record more health measurements during deterioration in the patient's health state. Therefore, AI solutions can use shorter time intervals to predict future decline in the trajectory. If the sampling is informative, we encourage researchers to develop models that consider the time intervals between the health measurements. For instance, 1 class of such models is point processes (eg, Hawkes processes). Here, a shorter time interval between health measurements makes further health measurements more likely and influences the expected risk score [23].

Figure 3 shows individual health recordings (eg, medical events, hospital visits, and laboratory data) in the form of dots. *Top:* an example patient trajectory where all medical events are equally spaced and thus there is a uniform time interval between the events. Here, the timing of the events is not informative of the current disease state. *Bottom:* an example patient trajectory that indicates a gradual decline in the disease state. Owing to this, additional health recordings are collected with higher frequency, which are thus informative about the disease state.

For objectives beyond risk scoring, we need other modeling approaches. When using risk scoring for *prescriptive purposes* (eg, treatment planning or dose finding), we encourage broader adoption of modeling strategies designed for decision making (eg, causal machine learning [37,68,69], Markov decision processes [70,71], dynamic treatment regimens [72,73], and policy learning [74]). There is a growing traction to extend many of these modeling strategies to handle longitudinal data, where health practitioners make treatment decisions over time.

Figure 3. Difference between noninformative and informative sampling.



For *subtyping*, one typically draws upon time-series clustering [26]. For this objective, the definition of a similarity metric is key [75,76]. One option is to set explicit rules for establishing patient trajectory similarities, for example, according to a specific condition such as heart disease [77] or a combination of clinical or phenotypic features [78]. Domain experts may be familiar with such approaches, but their definition may not perfectly integrate with AI algorithms and may thus require customization. Hence, an alternative is to view patient trajectory analysis from a methodological, *data-driven* angle. For example, when modeling the underlying temporal dynamics for performing data-driven clustering of longitudinal data, we can loosely group the approaches into (1) model-free approaches and (2) model-based approaches. In the model-free approaches, a similarity metric on sequential data is defined and then serves as input to conventional clustering algorithms. On the other hand, in the model-based approaches [79], a representation of the patient trajectory is learned, and the model parameters are then used for clustering (eg, mixture hidden Markov models). For future research, we find model-based approaches intriguing, as they no longer focus on raw observations but cluster the underlying disease dynamics (and, as such, can account for different latent disease states).

For *pathway discovery*, several descriptive approaches have emerged that allow for the discovery of subpatterns (ie, motifs) and common patient trajectories [76]. For example, Beck et al [33] analyzed disease pathways leading to septicemia in 110,000 patients. The study revealed prototypical pathways starting from 3 initial states (alcohol abuse, diabetes, and anemia) and established the trajectory-specific probability of sepsis mortality. This and similar studies reveal great potential to further our understanding of disease etiology and the possible means of changing disease trajectories [80,81]. Similarly, Oh et al [82] constructed patient trajectories consisting of specific health events (hyperlipidemia, hypertension, and impaired fasting glucose) and evaluated the probabilities of such trajectories (and their permutation) in increasing or decreasing the log odds of developing type 2 diabetes mellitus. Zhang and Padman [83] identified the most probable clinical trajectories from patients with chronic kidney disease by first grouping the patients and then fitting a first-order hidden Markov model to infer the most probable clinical pathways given sequences of multiple laboratory test observations and other patient characteristics. Huang et al [84] proposed a probabilistic model (based on latent Dirichlet allocation) to identify clinical pathway patterns from the event logs for patients with unstable angina and cancer. Dabek and Caban [85] offered another perspective by analyzing trajectories using automata (ie, deterministic and nondeterministic finite state automata) and using a grammar induction algorithm to identify common trajectories in neurology. Further approaches for pathway discovery are based on association rule mining [86] and functional dependencies mining [87].

Related to these objectives are models that adopt a structural lens to examine the causal mechanisms [88]. This would allow not only to understand *how* health measurements change over time but also *why*. The underlying AI algorithms are still under active research (eg, causal structure learning and neural causal discovery [89]). Here, it will be a rewarding direction for the future to develop more approaches that are tailored to longitudinal data.

Evaluation

Evaluations through randomized controlled trials are needed to confirm the effectiveness of AI-based analysis of patient trajectories in clinical practice. Recently, there have been such trials for traditional AI solutions that rely on snapshot data from a single or few time points [90]. However, similar trials for patient trajectory analysis are rare [90]. We expect significant value in conducting such trials and foresee challenges owing to the unique characteristics of patient trajectories. Foremost, evaluations through rigorous randomized controlled trials are a prerequisite to building trust in clinical practice, thereby expediting further AI-based trajectory analysis. However, evaluating such an AI solution might be a multiyear undertaking depending on the time window of the patient's trajectory. Thus, it is also essential to recognize that the evaluations are likely to involve a 2-step procedure. In the first step, trajectory data are collected to train the AI solution. In the second step, the previously trained AI solution is then deployed to analyze how the AI solution generalizes to new patient trajectories. When conducting such trials, it is crucial to acknowledge that the patient trajectories capture data from multiple time points and might thus be subject to an inherent domain shift (ie, where data distributions change over time, that is, over the patient journey) [91]. Such domain shifts might affect the performance of AI solutions [92], especially when the patient trajectories span a long time horizon. For example, AI-based predictions have been recently applied to patients with COVID-19 to compare the predicted health trajectory with the observed trajectory in a prospective study, finding that the performance of some risk scores decreased over time [93]. One reason was because of temporal domain shifts over time [94] as medical professionals learned about the emerging infectious disease and adapted their clinical routines over time, thus yielding different and, in particular, better outcomes than in the data used for training.

Furthermore, there is a need to ensure reproducibility of the AI-based analyses. Here, we consider 3 priorities. First, there is a need to develop a framework. For instance, the existing AI frameworks (such as scikit learn in Python) are designed for modeling static data. Conversely, more effort is necessary to define standardized building blocks for longitudinal data to effectively model the patient trajectories. Moreover, such frameworks should also involve tools for automation so that the disease models can be trained in a semiautomated manner (subsumed under the term *automated machine learning* [*AutoML*]). Although AutoML has become widespread for static data sets, only a few libraries are designed for time-series AutoML [95]. Here, we see enormous potential for future research at the interface to digital medicine. On the basis of our own experience, we expect such frameworks to play a critical role in achieving scalable and increased adoption of AI-based patient trajectory analysis in clinical settings. Second, future research should carefully assess and, if needed, revise the best practice guidelines for AI in medicine [96], so that they consider the longitudinal form of patient trajectories. For example, the reporting guidelines should involve information on whether the period between health measurements is informative (however, such information is not part of existing reporting guidelines for static patient data). Other examples could be the choice of the model (eg, whether latent dynamics were considered and why or which alternative architectures of recurrent neural networks were tested and eventually discarded), how the timing of the health care event was collected (eg, whether a time stamp was retrieved from a medical health record or whether this was a survey question referring to the *last 12 months* and the resulting uncertainty about the correct time), or the rationale behind how patient similarity was measured in subtyping. Finally, more data sets with patient trajectories should be made publicly available for benchmarking. Although data access is a common issue for AI research in medicine in general, the challenges are exacerbated in the context of patient journeys, where it is common to merge the health measurements from different sources (eg, from other health registries). So far, only a few longitudinal data sets are public (as compared with static data sets with patient data) [76]. Notable exceptions are large initiatives, such as the Healthcare Cost and Utilization Project, that offer longitudinal data sets for evaluating trajectory-based AI approaches [97].

Interpretations

To generate insights for clinical practice, it is often necessary that AI solutions overcome their black-box nature [98]. Here, we see enormous potential for new AI solutions that adhere to the needs of clinical practice with the objective of knowledge discovery.

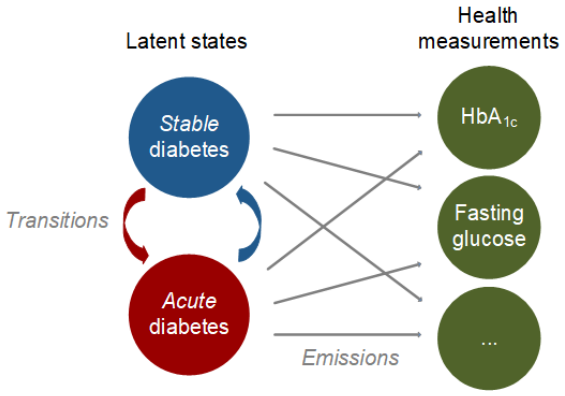
One approach is explainability, which aims to understand how a model arrives at a particular outcome [10,99]. However, AI explainability is typically developed in the context of static data and therefore, meaningful time-varying patterns from the course of a disease might not be revealed. For instance, SHAP values [99] inform which health measurements are used by the AI model and what values indicate risk. However, SHAP values cannot directly interpret the *dynamics* in health measurements (eg, whether there is an increase or decrease or large variability in health measurements, which would be needed to characterize changes in disease states over time). As a road map for research in digital medicine, we require techniques that interpret the temporal dynamics of disease progression, thereby being closely aligned with the demands of clinical practice. Here, digital medicine might find inspiration in other disciplines, for instance, financial technical analysis [100], where a systematic set of short-term movements of stock prices is used for interpretation. Similar patterns in patient trajectories could be inferred by researchers via short-term patterns (ie, trajectory markers) that characterize a disease course or a combination of short- and long-term motifs that help identify distinct disease states.

Another approach to generating insights is via interpretability, which builds upon inferences where the underlying logic is transparent. Interpretability often requires tailored modeling approaches. For static data, this is usually achieved through (penalized) linear regression or decision trees, whereas interpretability for longitudinal data is typically achieved through parsimonious models. Different strategies exist to obtain parsimonious models. For neural networks, there are techniques that tweak a neural network to provide a sparse one with similar performance (eg, enforcing feature sparsity through architecture design, modifying the objective function and the weight updating scheme [101], post hoc via reservoir computing or pruning, or a priori via cognitive networks [102]). Alternatively, one can draw upon structural formalizations (eg, dynamic fuzzy cognitive maps [78] to simulate patient trajectories) and probabilistic models (eg, Markov models, hidden Markov models, and Hawkes processes [23,25,63-67]).

Out of these modeling approaches, we expect hidden Markov models to be beneficial for interpretability, especially for risk scoring. The reason being hidden Markov models use latent variables to capture different disease phases in patient trajectories. These latent variables often have clinically relevant meanings and can thus be mapped onto existing clinical terminology (Figure 4). For instance, in diabetes mellitus, the latent states are characterized as *acute* and *stable* disease states [64] and thus are of clinical meaning. In addition, recent evidence suggests that interpretable models might also improve prediction performance [23]; however, more effort is needed to explore this further. In the future, we expect to see other modeling approaches that combine the strengths of hidden Markov models (for interpretability) with neural learning (for representation learning and capturing long-term dependencies in patient trajectories).

The health measurements are observable and thus obtained via standard data collection practices. In contrast, the latent states cannot be observed directly and, instead, are recovered from the health measurements. The latent states then describe different disease states in a patient trajectory (eg, *acute* vs *stable* disease states). During estimation, the latent states and health measurements are mathematically linked via components for both transition and emission probabilities.

Figure 4. Example of a hidden Markov model. HbA1c: hemoglobin A1c.



Implications for Digital Medicine

Analyzing patient trajectories using AI has multiple benefits. Dissecting the variability of disease pathways allows research to better understand both the disease etiology and disease course, facilitating a more extensive personalization of care. For instance, it enables the identification of short-term patterns predictive of future health states, which are then used during risk scoring. Similarly, patient trajectories capture the responsiveness of patients to treatments, and by leveraging this information in patient trajectories, AI solutions can guide treatment planning. In summary, AI-based trajectory analysis promises to strengthen the existing computational approaches to prevent, detect, diagnose, and treat diseases.

To address these challenges, we see particular importance in community building and in the development of computational patient trajectory tools that lower the barrier of entry. Community building will help to set a clear agenda and define an established terminology, bridging both practice and research in digital medicine. Here, we point to several valuable directions: (1) further effort will be needed to extend traditional clinical terminology (eg, cohort building and patient similarity) to AI-based trajectory analysis, thereby facilitating communication between the experts in AI and medicine; (2) it is essential to build communities by connecting different actors from regulation, law, data science, and medicine, as this will eventually be a prerequisite for deploying AI solutions in medical practice; and (3) such communities may promote data exchange, thus allowing for more extensive benchmarking of AI solutions. Similarly, we also suggest hosting leaderboard competitions as conducted in other fields (eg, the SemEval benchmark competitions in natural language processing). Leaderboard competitions will eventually help to identify robust AI solutions and thus to condense best practices during modeling.

Acknowledgments

SF received funding from the Swiss National Science Foundation (grant 186932). AA and MK were funded by grant 201184 from the Swiss National Science Foundation. The opinions expressed in this manuscript are those of the authors' and do not necessarily reflect those of the Novartis Institutes for Biomedical Research.

Conflicts of Interest

MR declares employment with the Novartis Institutes for Biomedical Research, Switzerland.

References

1. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, et al. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nat Commun 2014 Jun 24;5(1):4022 [FREE Full text] [doi: 10.1038/ncomms5022] [Medline: 24959948]
2. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
3. 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]
4. Woldaregay AZ, Årsand E, Botsis T, Albers D, Mamykina L, Hartvigsen G. Data-driven blood glucose pattern classification and anomalies detection: machine-learning applications in type 1 diabetes. J Med Internet Res 2019 May 01;21(5):e11030 [FREE Full text] [doi: 10.2196/11030] [Medline: 31042157]
5. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med 2018 Jan 15;1(1):20172. [doi: 10.1038/s41746-017-0002-4]

6. Porumb M, Stranges S, Pescapè A, Pecchia L. Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG. Sci Rep 2020 Jan 13;10(1):170 [FREE Full text] [doi: 10.1038/s41598-019-56927-5] [Medline: 31932608]
7. 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]
8. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. Sci Rep 2018 Apr 17;8(1):6085 [FREE Full text] [doi: 10.1038/s41598-018-24271-9] [Medline: 29666385]
9. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. J Biomed Inform 2018 Jul;83:112-134 [FREE Full text] [doi: 10.1016/j.jbi.2018.04.007] [Medline: 29879470]
10. Thorsen-Meyer H, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Health 2020 Apr;2(4):179-191. [doi: 10.1016/s2589-7500(20)30018-2]
11. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data 2019 Jun 17;6(1):96 [FREE Full text] [doi: 10.1038/s41597-019-0103-9] [Medline: 31209213]
12. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. Sci Rep 2019 Jan 24;9(1):717 [FREE Full text] [doi: 10.1038/s41598-018-36745-x] [Medline: 30679510]
13. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020 Sep 24;3(1):126 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 34556791]
14. Pencina MJ, Goldstein BA, D'Agostino RB. Prediction models — development, evaluation, and clinical application. N Engl J Med 2020 Apr 23;382(17):1583-1586. [doi: 10.1056/nejmp2000589]
15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
16. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/nejmra1814259]
17. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
18. Tomašev N, Harris N, Baur S, Mottram A, Glorot X, Rae JW, et al. Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. Nat Protoc 2021 Jun 05;16(6):2765-2787. [doi: 10.1038/s41596-021-00513-5] [Medline: 33953393]
19. Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ Digit Med 2021 Jan 08;4(1):5 [FREE Full text] [doi: 10.1038/s41746-020-00376-2] [Medline: 33420381]
20. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. Br Med J 2020 Sep 09;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
21. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. Br Med J 2020 Sep 09;370:m3210 [FREE Full text] [doi: 10.1136/bmj.m3210] [Medline: 32907797]
22. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan 7;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
23. Alaa A, Hu S, Schaar M. Learning from clinical judgments: Semi-Markov-modulated marked Hawkes processes for risk prognosis. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia p. 60-69 URL: https://proceedings.mlr.press/v70/alaa17a.html
24. Ghassemi M, Naumann T, Schulam P, Beam A, Chen I, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191-200 [FREE Full text] [Medline: 32477638]
25. Oezyurt Y, Kraus M, Hatt T, Feuerriegel S. AttDMM: an attentive deep Markov model for risk scoring in intensive care units. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021 Presented at: KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; August 14 - 18, 2021; Virtual Event Singapore p. 3452-3462. [doi: 10.1145/3447548.3467143]
26. Schulam P, Wigley F, Saria S. Clustering longitudinal clinical marker trajectories from electronic health data: applications to phenotyping and endotype discovery. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015 Presented at: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence; January 25 - 30, 2015; Austin Texas p. 2956-2964. [doi: 10.5555/2888116.2888128]
27. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
28. Lipton Z, Kale D, Elkan C, Wetzel R. Learning to diagnose with LSTM recurrent neural networks. In: Proceedings of the International Conference on Learning Representations (ICLR). 2015 Presented at: International Conference on Learning

Representations (ICLR); May 7-9, 2015; San Diego, CA, USA URL: https://scholar.google.fr/citations?view_op=view_citation&hl=en&user=MN9Kfg8AAAAJ&citation_for_view=MN9Kfg8AAAAJ:Tyk-4Ss8FVUC
29. Zhang D, Yin C, Hunold KM, Jiang X, Caterino JM, Zhang P. An interpretable deep-learning model for early prediction of sepsis in the emergency department. Patterns (N Y) 2021 Feb 12;2(2):100196 [FREE Full text] [doi: 10.1016/j.patter.2020.100196] [Medline: 33659912]
30. Li Y, Rao S, Solares JR, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: Transformer for electronic health records. Sci Rep 2020 Apr 28;10(1):7155 [FREE Full text] [doi: 10.1038/s41598-020-62922-y] [Medline: 32346050]
31. Choi Y, Chiu CY, Sontag D. Learning low-dimensional representations of medical concepts. AMIA Jt Summits Transl Sci Proc 2016;2016:41-50 [FREE Full text] [Medline: 27570647]
32. Alaa AM, Yoon J, Hu S, van der Schaar M. Personalized risk scoring for critical care prognosis using mixtures of Gaussian processes. IEEE Trans Biomed Eng 2018 Jan;65(1):207-218. [doi: 10.1109/tbme.2017.2698602]
33. Beck MK, Jensen AB, Nielsen AB, Perner A, Moseley PL, Brunak S. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. Sci Rep 2016 Nov 04;6(1):36624 [FREE Full text] [doi: 10.1038/srep36624] [Medline: 27812043]
34. Hirano S, Tsumoto S. Mining similar temporal patterns in long time-series data and its application to medicine. In: Proceedingsd of the IEEE International Conference on Data Mining. 2002 Presented at: IEEE International Conference on Data Mining; Dec. 9-12, 2002; Maebashi City, Japan. [doi: 10.1109/icdm.2002.1183906]
35. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug 31;572(7767):116-119 [FREE Full text] [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
36. Ravizza S, Huschto T, Adamov A, Böhm L, Büsser A, Flöther FF, et al. Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. Nat Med 2019 Jan 7;25(1):57-59. [doi: 10.1038/s41591-018-0239-8] [Medline: 30617317]
37. Schulam P, Saria S. Reliable decision support using counterfactual models. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: 31st International Conference on Neural Information Processing Systems; December 4 - 9, 2017; Long Beach California USA p. 1696-1706. [doi: 10.5555/3294771.3294933]
38. Lim B, Alaa A, van der Schaar M. Forecasting treatment responses over time using recurrent marginal structural networks. In: Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018). 2018 Presented at: 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); 2018; Montréal, Canada URL: https://proceedings.neurips.cc/paper/2018/file/56e6a93212e4482d99c84a639d254b67-Paper.pdf
39. Liu R, Yin C, Zhang P. Estimating individual treatment effects with time-varying confounders. In: Proceedings of the IEEE International Conference on Data Mining (ICDM). 2020 Presented at: IEEE International Conference on Data Mining (ICDM); Nov. 17-20, 2020; Sorrento, Italy. [doi: 10.1109/icdm50108.2020.00047]
40. Hatt T, Feuerriegel S. Sequential deconfounding for causal inference with unobserved confounders. arXiv. 2021. URL: https://arxiv.org/abs/2104.09323 [accessed 2021-11-27]
41. Bica I, Alaa A, Jordon J, van der Schaar M. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In: Proceedings of the International Conference on Learning Representations (ICLR). 2019 Presented at: International Conference on Learning Representations (ICLR); May 6-9, 2019; New Orleans, LA, USA URL: https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=mnU3HpcAAAAJ&citation_for_view=mnU3HpcAAAAJ:9yKSN-GCB0IC
42. Kuzmanovic M, Hatt T, Feuerriegel S. Deconfounding Temporal Autoencoder: estimating treatment effects over time using noisy proxies. In: Proceedings of the Machine Learning for Health (ML4H) Workshop 2021. 2021 Presented at: Machine Learning for Health (ML4H) Workshop 2021; December 4, 2021; Online (Upcoming) URL: https://www.research-collection.ethz.ch/handle/20.500.11850/512184
43. Izenman AJ. Introduction to manifold learning. WIREs Comp Stat 2012 Jul 16;4(5):439-446. [doi: 10.1002/wics.1222]
44. Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci U S A 2016 Jul 05;113(27):7329-7336 [FREE Full text] [doi: 10.1073/pnas.1510502113] [Medline: 27274072]
45. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020 Sep 14;3(1):119 [FREE Full text] [doi: 10.1038/s41746-020-00323-1] [Medline: 34518641]
46. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep 2020 Jul 28;10(1):12598 [FREE Full text] [doi: 10.1038/s41598-020-69250-1] [Medline: 32724046]
47. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell 2020 Jun 08;2(6):305-311. [doi: 10.1038/s42256-020-0186-1]
48. McMahan B, Moore E, Ramage D, Hampson S, y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). 2017 Presented at: 20th International Conference on Artificial Intelligence and Statistics (AISTATS); April 20-22, 2017; Fort Lauderdale, FL, USA.

49. Li T, Sahu A, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. MLSys. 2020. URL: https://arxiv.org/abs/1812.06127 [accessed 2021-11-27]
50. Konečný J, McMahan H, Ramage D, Richtárik P. Federated optimization: distributed machine learning for on-device intelligence. arXiv. 2016. URL: https://arxiv.org/abs/1610.02527 [accessed 2021-11-27]
51. Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, Ienca M, Fellay J, Vayena E, et al. Revolutionizing medical data sharing using advanced privacy-enhancing technologies: technical, legal, and ethical synthesis. J Med Internet Res 2021 Feb 25;23(2):e25120 [FREE Full text] [doi: 10.2196/25120] [Medline: 33629963]
52. Liu D, Dligach D, Miller T. Two-stage federated phenotyping and patient representation learning. In: Proceedings of the 18th BioNLP Workshop and Shared Task. 2019 Presented at: 18th BioNLP Workshop and Shared Task; August, 2019; Florence, Italy p. 283-291. [doi: 10.18653/v1/w19-5030]
53. Liu D, Miller T, Sayeed R, Mandl KD. FADL: Federated-autonomous deep learning for distributed electronic health record. In: Proceedings of the NIPS Machine Learning for Health (ML4H) Workshop. 2018 Presented at: NIPS Machine Learning for Health (ML4H) Workshop; December 08, 2018; Montreal, Canada URL: https://arxiv.org/abs/1811.11400
54. Huang L, Yin Y, Fu Z, Zhang S, Deng H, Liu D. LoAdaBoost: Loss-based AdaBoost federated machine learning with reduced computational complexity on IID and non-IID intensive care data. PLoS One 2020 Apr 17;15(4):e0230706 [FREE Full text] [doi: 10.1371/journal.pone.0230706] [Medline: 32302316]
55. Huang L, Shea AL, Qian H, Masurkar A, Deng H, Liu D. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. J Biomed Inform 2019 Nov;99:103291 [FREE Full text] [doi: 10.1016/j.jbi.2019.103291] [Medline: 31560949]
56. Lee J, Sun J, Wang F, Wang S, Jun C, Jiang X. Privacy-preserving patient similarity learning in a federated environment: development and analysis. JMIR Med Inform 2018 Apr 13;6(2):e20 [FREE Full text] [doi: 10.2196/medinform.7744] [Medline: 29653917]
57. Li W, Milletarì F, Xu D, Rieke N, Hancox J, Zhu W, et al. Privacy-preserving federated brain tumour segmentation. In: Proceedings of the International Workshop on Machine Learning in Medical Imaging. 2019 Presented at: International Workshop on Machine Learning in Medical Imaging; October 13, 2019; Shenzhen, China p. 133-141. [doi: 10.1007/978-3-030-32692-0_16]
58. Gentry C. Computing arbitrary functions of encrypted data. Commun ACM 2010 Mar;53(3):97-105. [doi: 10.1145/1666420.1666444]
59. Choudhury O, Gkoulalas-Divanis A, Salonidis T, Sylla I, Park Y, Hsu G, et al. Anonymizing data for privacy-preserving federated learning. In: Proceedings of the European Conference on Artificial Intelligence (ECAI). 2020 Presented at: European Conference on Artificial Intelligence (ECAI); 2020; Santiago de Compostela, Spain URL: https://arxiv.org/abs/2002.09096
60. Kholod I, Yanaki E, Fomichev D, Shalugin E, Novikova E, Filippov E, et al. Open-source federated learning frameworks for IoT: a comparative review and analysis. Sensors 2020 Dec 29;21(1):167 [FREE Full text] [doi: 10.3390/s21010167] [Medline: 33383803]
61. Beutel D, Topal T, Mathur A, Qiu X, Parcollet T, Gusmão PD, et al. Flower: a friendly federated learning research framework. arXiv. 2020. URL: https://arxiv.org/abs/2007.14390 [accessed 2021-11-27]
62. Brouwer ED, Simm J, Arany A, Moreau Y. Deep ensemble tensor factorization for longitudinal patient trajectories classification. arXiv. 2018. URL: https://arxiv.org/abs/1811.10501 [accessed 2021-11-26]
63. Naumzik C, Feuerriegel S, Nielsen A. Data-driven dynamic treatment planning for chronic diseases. Working Paper 2022:1-30 [FREE Full text]
64. Maag B, Feuerriegel S, Kraus M, Saar-Tsechansky M, Züger T. Modeling longitudinal dynamics of comorbidities. In: Proceedings of the Conference on Health, Inference, and Learning (CHIL). 2021 Presented at: Conference on Health, Inference, and Learning (CHIL); April 8 - 10, 2021; Virtual. [doi: 10.1145/3450439.3451871]
65. Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW. Hidden Markov models for alcoholism treatment trial data. Ann Appl Stat 2010 Mar 1;4(1):366-395. [doi: 10.1214/09-aoas282]
66. Scott SL, James GM, Sugar CA. Hidden Markov models for longitudinal comparisons. J Am Stat Assoc 2005;100(470):359-369. [doi: 10.1198/016214504000001592]
67. Bartolomeo N, Trerotoli P, Serio G. Progression of liver cirrhosis to HCC: an application of hidden Markov model. BMC Med Res Methodol 2011;11(1):38 [FREE Full text] [doi: 10.1186/1471-2288-11-38] [Medline: 21457586]
68. Daniel R, Cousens S, De Stavola B, Kenward MG, Sterne JA. Methods for dealing with time-dependent confounding. Stat Med 2013;32(9):1584-1618. [doi: 10.1002/sim.5686] [Medline: 23208861]
69. Hatt T, Feuerriegel S. Estimating average treatment effects via orthogonal regularization. In: Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM). 2021 Presented at: ACM International Conference on Information and Knowledge Management (CIKM); November 1 - 5, 2021; Virtual. [doi: 10.1145/3459637.3482339]
70. Schaefer A, Bailey M, Shechter S, Roberts M. Modeling medical treatment using Markov decision processes. Oper Res Health Care 2005;70:593-612. [doi: 10.1007/1-4020-8066-2_23]

71. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov 22;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
72. Murphy SA, van der Laan MJ, Robins JM, CPPRG. Marginal mean models for dynamic regimes. J Am Stat Assoc 2001 Dec 01;96(456):1410-1423 [FREE Full text] [doi: 10.1198/016214501753382327] [Medline: 20019887]
73. Murphy S. Optimal dynamic treatment regimes. J Royal Stat Soc : Series B (Stat Methodol) 2003;65(2):331-355. [doi: 10.1111/1467-9868.00389]
74. Liao P, Klasnja P, Murphy S. Off-policy estimation of long-term average outcomes with applications to mobile health. J Am Stat Assoc 2021 Oct 01;116(533):382-391. [doi: 10.1080/01621459.2020.1807993] [Medline: 33814653]
75. Brown S. Patient similarity: emerging concepts in systems and precision medicine. Front Physiol 2016 Nov 24;7:561 [FREE Full text] [doi: 10.3389/fphys.2016.00561] [Medline: 27932992]
76. Allam A, Dittberner M, Sintsova A, Brodbeck D, Krauthammer M. Patient similarity analysis with longitudinal health data. arXiv. 2020. URL: https://arxiv.org/abs/2005.06630 [accessed 2021-11-26]
77. Carrasco-Ribelles LA, Pardo-Mas JR, Tortajada S, Sáez C, Valdivieso B, García-Gómez JM. Predicting morbidity by local similarities in multi-scale patient trajectories. J Biomed Inform 2021 Aug;120:103837. [doi: 10.1016/j.jbi.2021.103837] [Medline: 34119690]
78. Giabbanelli PJ, Torsney-Weir T, Mago VK. A fuzzy cognitive map of the psychosocial determinants of obesity. Appl Soft Comput 2012 Dec;12(12):3711-3724. [doi: 10.1016/j.asoc.2012.02.006]
79. Ghassempour S, Girosi F, Maeder A. Clustering multivariate time series using hidden Markov models. Int J Environ Res Public Health 2014 Mar 06;11(3):2741-2763 [FREE Full text] [doi: 10.3390/ijerph110302741] [Medline: 24662996]
80. Westergaard D, Moseley P, Sørup FKH, Baldi P, Brunak S. Population-wide analysis of differences in disease progression patterns in men and women. Nat Commun 2019 Feb 08;10(1):666 [FREE Full text] [doi: 10.1038/s41467-019-08475-9] [Medline: 30737381]
81. Lademann M, Lademann M, Jensen AB, Brunak S. Incorporating symptom data in longitudinal disease trajectories for more detailed patient stratification. Int J Med Inform 2019 Sep;129:107-113 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.003] [Medline: 31445244]
82. Oh W, Kim E, Castro MR, Caraballo PJ, Kumar V, Steinbach MS, et al. Type 2 diabetes mellitus trajectories and associated risks. Big Data 2016;4(1):25-30 [FREE Full text] [doi: 10.1089/big.2015.0029] [Medline: 27158565]
83. Zhang Y, Padman R. Innovations in chronic care delivery using data-driven clinical pathways. Am J Manag Care 2015 Dec 01;21(12):661-668 [FREE Full text] [Medline: 26760429]
84. Huang Z, Dong W, Ji L, Gan C, Lu X, Duan H. Discovery of clinical pathway patterns from event logs using probabilistic topic models. J Biomed Inform 2014 Feb;47:39-57 [FREE Full text] [doi: 10.1016/j.jbi.2013.09.003] [Medline: 24076435]
85. Dabek F, Caban J. A grammar-based approach to model the patient's clinical trajectory after a mild traumatic brain injury. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2015 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 9 - 12, 2015; Washington, DC. [doi: 10.1109/bibm.2015.7359775]
86. Li X, Liu H, Mei J, Yu Y, Xie G. Mining temporal and data constraints associated with outcomes for care pathways. Stud Health Technol Inform 2015;216:711-715. [Medline: 26262144]
87. Combi C, Mantovani M, Sala P. Discovering quantitative temporal functional dependencies on clinical data. In: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); August 23 - 26, 2017; Park City, UT. [doi: 10.1109/ichi.2017.80]
88. Giabbanelli PJ. Analyzing the complexity of behavioural factors influencing weight in adults. In: Giabbanelli PJ, Mago V, Papageorgiou E, editors. Advanced Data Analytics in Health. Cham, Switzerland: Springer; 2018:163-181.
89. Scherrer N, Bilaniuk O, Annadani Y, Goyal A, Schwab P, Schölkopf B, et al. Learning neural causal models with active interventions. arXiv. 2021. URL: https://arxiv.org/abs/2109.02429 [accessed 2021-11-26]
90. Topol EJ. Welcoming new guidelines for AI clinical research. Nat Med 2020 Sep 09;26(9):1318-1320. [doi: 10.1038/s41591-020-1042-x] [Medline: 32908274]
91. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. Commun Med 2021;1(1):25. [doi: 10.1038/s43856-021-00028-w] [Medline: 34522916]
92. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med 2021 Jul 15;385(3):283-286. [doi: 10.1056/nejmc2104626]
93. Arık SO, Shor J, Sinha R, Yoon J, Ledsam JR, Le LT, et al. A prospective evaluation of AI-augmented epidemiology to forecast COVID-19 in the USA and Japan. NPJ Digit Med 2021 Oct 08;4(1):146 [FREE Full text] [doi: 10.1038/s41746-021-00511-7] [Medline: 34625656]
94. Roland T, Boeck C, Tschoellitsch T, Maletzky A, Hochreiter S, Meier J, et al. Machine learning based COVID-19 diagnosis from blood tests with robustness to domain shifts. medRxiv. 2021. URL: https://www.medrxiv.org/content/10.1101/2021.04.06.21254997v1 [accessed 2021-11-26]
95. Jarrett D, Yoon J, Bica I, Qian Z, Ercole A, van der Schaar M. Clairvoyance: a unified, end-to-end AutoML pipeline for medical time series. In: Proceedings of the International Conference on Learning Representations (ICLR). 2021 Presented

at: International Conference on Learning Representations (ICLR); May 3 - 7, 2021; Vienna, Austria URL: https://openreview.net/forum?id=xnC8YwKUE3k
96. Kakarmath S, Esteva A, Arnaout R, Harvey H, Kumar S, Muse E, et al. Best practices for authors of healthcare-related artificial intelligence manuscripts. NPJ Digit Med 2020 Oct 16;3(1):134 [FREE Full text] [doi: 10.1038/s41746-020-00336-w] [Medline: 33083569]
97. Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. Sci Rep 2019 Jun 26;9(1):9277 [FREE Full text] [doi: 10.1038/s41598-019-45685-z] [Medline: 31243311]
98. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018 Mar 14;1(1):5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
99. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018 Oct 10;2(10):749-760 [FREE Full text] [doi: 10.1038/s41551-018-0304-0] [Medline: 31001455]
100. Kirkpatrick C, Dahlquist J. Technical Analysis: The Complete Resource for Financial Market Technicians. Upper Saddle River, NJ: Financial Times Press; 2006.
101. Lemhadri I, Ruan F, Abraham L, Tibshirani R. LassoNet: a neural network with feature sparsity. J Machin Learn Res 2021;22(127):1-29 [FREE Full text]
102. Nápoles G, Jastrzębska A, Salgueiro Y. Pattern classification with evolving long-term cognitive networks. Inform Sci 2021 Feb;548:461-478. [doi: 10.1016/j.ins.2020.08.058]

Abbreviations

AI: artificial intelligence
AutoML: automated machine learning
ICU: intensive care unit
RNN: recurrent neural network

Edited by R Kukafka; submitted 21.04.21; peer-reviewed by Z Ren, P Giabbanelli, E Mahmoudi; comments to author 27.05.21; revised version received 26.07.21; accepted 29.10.21; published 03.12.21.

Please cite as:
Allam A, Feuerriegel S, Rebhan M, Krauthammer M
Analyzing Patient Trajectories With Artificial Intelligence
J Med Internet Res 2021;23(12):e29812
URL: https://www.jmir.org/2021/12/e29812
doi:10.2196/29812
PMID:34870606

Viewpoint

Examining TikTok's Potential for Community-Engaged Digital Knowledge Mobilization With Equity-Seeking Groups

Kinnon Ross MacKinnon[1], MSW, PhD; Hannah Kia[2*], MSW, PhD; Ashley Lacombe-Duncan[3*], MSW, PhD

[1]School of Social Work, York University, Toronto, ON, Canada

[2]School of Social Work, The University of British Columbia, Vancouver, BC, Canada

[3]School of Social Work, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Kinnon Ross MacKinnon, MSW, PhD
School of Social Work
York University
4700 Keele Avenue
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100 ext 66333
Email: kinnonmk@yorku.ca

Abstract

Social media is increasingly being leveraged by researchers to engage in public debates and rapidly disseminate research results to health care providers, health care users, policy makers, educators, and the general public. This paper contributes to the growing literature on the use of social media for digital knowledge mobilization, drawing particular attention to TikTok and its unique potential for collaborative knowledge mobilization with underserved communities who experience barriers to health care and health inequities (eg, equity-seeking groups). Setting the TikTok platform apart from other social media are the unique audiovisual video editing tools, together with an impactful algorithm, that make knowledge dissemination and exchange with large global audiences possible. As an example, we will discuss digital knowledge mobilization with trans and nonbinary (trans) communities, a population that experiences barriers to health care and is engaged in significant peer-to-peer health information sharing on the web. To demonstrate, analytics data from 13 selected TikTok videos on the topic of research on gender-affirming medicine (eg, hormonal therapy and surgeries) are presented to illustrate how knowledge is disseminated within the trans community via TikTok. Considerations for researchers planning to use TikTok for digital knowledge mobilization and other related community engagement with equity-seeking groups are also discussed. These include the limitations of TikTok analytics data for measuring knowledge mobilization, population-specific concerns related to community safety on social media, the spread of disinformation, barriers to internet access, and commercialization and intellectual property issues. This paper concludes that TikTok is an innovative social media platform that presents possibilities for achieving transformative, community-engaged knowledge mobilization among researchers, underserved health care users, and their health care providers, all of whom are necessary to achieve better health care and population health outcomes.

(J Med Internet Res 2021;23(12):e30315) doi:10.2196/30315

KEYWORDS

trans; nonbinary; marginalized communities; gender-affirming care; digital health; community-engaged research; knowledge mobilization; mobile phone

Introduction

Background

Social media is increasingly being leveraged by researchers to engage in scholarly debates and rapidly disseminate research results. Rapid dissemination strategies can facilitate the uptake of life-saving information, fill gaps in critical information, and correct misinformation, as has been documented throughout the COVID-19 pandemic [1]. However, social media also provides a platform for challenging disseminated knowledge that is produced about communities without appropriate community engagement and critical attention to ethical considerations. As an example, significant controversy erupted in 2019 when a researcher published a controversial study that presented a medically unsubstantiated term, *rapid-onset gender dysphoria*,

arguing that trans and nonbinary (trans) identity is a *social and peer contagion* [2]. In response, several trans-identified community-engaged researchers challenged these findings on Twitter to draw attention to the study's methodological flaws and its potential harm to trans-people [3-5]. Scientific critiques of this study were later published by researchers in peer-reviewed journals and again shared through social media and web-based news media channels, leading to corrections made to the original article [6,7]. This is only one striking example demonstrating the transformative potential of community-engaged digital advocacy.

This paper extends the existing research on the use of social media for knowledge mobilization and collaborative knowledge exchange, focusing explicitly on TikTok. It is among the first academic papers to draw attention to this fast-growing social media platform and its potential use as a digital knowledge mobilization tool. In so doing, the article highlights (1) how TikTok functions, including its unique algorithm, which creates the opportunity to reach larger international audiences in comparison with other social media platforms and relatedly, (2) TikTok's potential for audiovisual knowledge exchange with communities who are active TikTok users—also referred to as *creators* on the platform. As an example, we specifically highlight the possibilities for knowledge mobilization with trans communities, drawing from the first author's (KRM) TikTok analytics data. Some of TikTok's anticipated pitfalls and limitations have also been discussed to inform researchers. This viewpoint study is also of practical use to researchers engaging with an expansive number of web-based communities on TikTok, such as youth, and more [8]. Although important, in-depth considerations of the intersections among social media technologies, ethics, and society are out of scope. For interested readers, these discussions may be sought out within science and technology studies and other interdisciplinary literature [9-11].

Knowledge Mobilization and Health Equity

Knowledge mobilization is of primary concern to researchers whose study findings have tremendous consequences on the health equity of underserved, minoritized populations, who are also termed *equity-seeking groups*. For example, trans individuals constitute a growing population whose shared experiences and health care needs have garnered significant attention in recent years. Trans persons identify with a different gender than was assigned at birth, with many undergoing a gender transition to align their physical characteristics with their internally felt gender identity [12]. Gender transitions may include the use of a new name and gender pronoun and/or changes to one's physical appearance (eg, clothing and hairstyle); updating some or all of one's legal identity documentation (eg, passport and driver's license); or accessing gender-affirming hormonal therapy or surgical procedures, among other social, legal, and medical processes [12]. Although not all trans people seek gender-affirming medicine (eg, hormones and surgeries) to transition, improving access to these medical treatments is of great concern to trans communities and their health care providers. Gender-affirming medicine is associated with improvements in trans people's mental health outcomes [13,14], and in some jurisdictions, these treatments are a prerequisite for changing legal sex designations [15]. Recently, in the United States and in Britain, there have been numerous legislative attempts to limit young trans people's access to gender-affirming medicine [16,17], despite evidence that such a law could increase mental health concerns such as suicide risk [13,17]. Compounding these health inequities, health care providers rarely receive sufficient education or training to provide care that is sensitive to, or tailored to the needs of, trans health care users [18,19]. The fear of nonaffirming or discriminatory and/or incompetent health care can contribute to delayed health care among trans persons and teaching their own health care providers, which has been shown to increase the odds of delaying needed medical care [20,21]. At the same time, this may explain much of the intracommunity, peer-to-peer sharing of health-related content via web-based platforms [22]. In this paper, we discuss why and how trans people's health care inequities can be addressed by bolstering digital knowledge mobilization and community-engaged knowledge exchange.

Community-engaged research broadly refers to various processes and practices involving collaborative partnerships with communities and/or people who have been historically underrepresented within, and poorly served by, the academic knowledge production process (eg, trans communities). An aspiration of community-engaged research is to form a symbiotic relationship whereby communities, for instance, health care users and researchers, work together with a common goal such as improving access to care [23]. Knowledge mobilization and collaborative knowledge exchange are integral components of community-engaged research. Knowledge mobilization is an umbrella term referring to the collaborative flow of research knowledge including "knowledge synthesis, dissemination, transfer, exchange, and cocreation or coproduction by researchers and knowledge users" [24]. Knowledge mobilization activities are associated with a faster translation of research findings to achieve transformation of publicly held beliefs, changing provider practices, inciting policy change, and exchanging or coconstructing knowledge together with key stakeholders or knowledge users.

Social Media, Knowledge Mobilization, and Exchange

The use of social media as a strategy for disseminating research findings beyond academic contexts has been previously established. Social media provides opportunities for knowledge to be publicly discussed and debated almost instantaneously by researchers, together with knowledge users such as educators, health care providers, health care users, and the general population. Croatian medical researchers have used Facebook to translate evidence-based medicine and high-quality health information to laypeople, health professionals, and journalists alike [25]. A 2019 scoping review identified a sharp rise in the use of Facebook, Twitter, and podcasts to translate knowledge between physicians and their trainees since 1996 [26]. In total, this scoping review revealed 12 different social media platforms used for knowledge mobilization. TikTok was not included in this list (likely because of the literature search parameters including studies published between 1990 and 2018). In another recent example, web-based discussions of hydroxychloroquine and remdesivir enabled public interdisciplinary medical discussions and the discrediting of false claims [27]. These constitute communicative spaces where diverging "viewpoints

have the potential to circulate and be negotiated" [28]. TikTok has also been noted as a platform through which credible COVID-19 public health messaging circulates [29], for instance, the collective responsibility to keep communities healthy by practicing social distancing [30]. A literature review of young people's use of digital technology highlighted that lesbian, gay, bisexual, trans, and intersex youth value storytelling about identity and concluded that there is a need to further establish how social media platforms influence young people's peer-to-peer sharing of health information [31]. Trans people, in particular, use social media as a *vehicle of transformation* to artistically document their gender transitions through a process of coproduction and storytelling with other trans people [32], with trans women engaging in extensive community-based digital advocacy on the web through the hashtag *#GirlsLikeUs* [33].

Increasingly, knowledge exchange and coconstruction of knowledge occur on social media platforms. Before the COVID-19 pandemic, much of researchers' community engagement work and knowledge mobilization occurred in person. However, due in part to widespread social distancing and lockdown policies intended to mitigate COVID-19 transmission, social media use has accelerated over the past year, as these technologies enable people to stay connected with family, friends, and community [34]. For instance, TikTok use exploded in 2020, with it being one of the most frequently downloaded smartphone apps [35]. Social media platforms now serve as replacement access points for communities and other social connections. For this reason, community-engaged researchers must continue to develop digital knowledge mobilization strategies to join communities in which they are organizing. At the present moment, trans communities are primarily connected on the web.

TikTok: A Social Media Platform for Short, Creative Knowledge Mobilization Videos

A subsidiary of the Beijing technology company ByteDance, TikTok currently sits among the world's most popular smartphone apps [35]. According to TikTok's *Our Mission* section, "TikTok is the leading destination for short-form mobile video. Our mission is to inspire creativity and bring joy" [35]. The TikTok platform enables users to create, share, and otherwise digitally engage with short videos (<3 minutes), which are simply referred to by users as *tiktoks*. Approximately 800 million people are registered as TikTok account holders worldwide [36]. In 2020, the proportion of global internet users aged 16 to 68 years engaging with TikTok was estimated to be approximately 18% [37]. Although userbase demographics include significant diversity in terms of gender, ethno-racial, cultural, and other characteristics, approximately 42% of TikTok users are aged 18 to 24 years [29]. Importantly, people >30 years are increasingly joining TikTok, including health care providers, researchers, and educators [38].

Integrating music, text, and video together, tiktoks aim to be audiovisually entertaining, and the platform itself provides creators with several built-in tools useful for engaging in digital knowledge mobilization and the collaborative exchange of ideas. Not unlike other social media platforms, such as Twitter, the TikTok video comment section enables text-based responses. However, TikTok users are also provided with the option of creating a video response to text comments, which makes more nuanced and in-depth audio-video discussions possible. Using the video-response-to-comment option, a creator can produce an entirely new tiktok, which features the original comment in a speech balloon. The 2 additional popular TikTok video editing options are the *stitch* and the *duet*. Creating a stitch video involves responding to another user's video by inserting—quite literally stitching—a segment of the original video to trigger the start of a new tiktok, establishing a public knowledge exchange between 2 users within the same short video. A duet serves a similar purpose; however, a duet typically amplifies and preserves the original video's content while at the same time offering notes of endorsement or critiques toward the original information shared. In many cases with duets or stitches, the original tiktok creator explicitly requested other users to stitch or duet the video to answer a question or otherwise engage in a larger public discussion. Taken together, these video creation options bolster a creative, audio-video digital knowledge exchange environment through which TikTok users engage with one another's videos. These in-house features of the TikTok app are especially conducive to community-engaged knowledge exchange, as they involve direct, audiovisual communication with other TikTok users.

Those who have experience with opening social media accounts on other platforms may appreciate that one of the first hurdles is gaining an audience (eg, subscribers or *followers*) with whom to exchange information surrounding particular topics. Contrasting with other social media apps built around individual subscribers, TikTok connects an international user base through a unique algorithm. Each TikTok user has a personalized *For You* page, which is fine-tuned by the TikTok algorithm, reflecting each user's individual interests and engagement on the platform. The algorithm is sensitive to the user's TikTok activity, including past videos commented on, liked, or shared, as well as the hashtags these videos apply. Not unlike other social media, it is primarily through all of these platform-based activities that a user's *For You* page is curated by the algorithm [39]. The TikTok platform may ameliorate this barrier of finding an audience to engage with, given that when a video's content resonates with other platform users, it may gain the attention of a large global audience in relatively little time as it is shown to other users [38]. TikTok videos shared, stitched, and duetted extensively by other users also tend to attract larger audiences, and the algorithm mirrors this activity by sending popular videos to be viewed by an increasing number of users, thereby compounding a trending video's attention.

To demonstrate, Basch et al [40] sampled 100 popular, trending English-language tiktoks that applied the hashtag *#WearAMask* and compared these with 32 videos posted by the World Health Organization (WHO), which were also related to mask wearing. The metadata of each of the 132 sampled videos were collected and analyzed. Although the ratio of trending tiktoks to WHO videos was approximately 3:1, TikTok videos were viewed 500 million times versus the WHO's 57 million views. The *#WearAMask* TikTok campaign received approximately 10 times the views, indicating *widespread reach* and TikTok's

transformative, community-engaged potential for communicating health information. Eghtesadi and Florea [41] also observed that TikTok is an important social media platform that could be used to share medical research with physicians and the general public. Useful for public knowledge exchange on topics of community concern, TikTok could help to achieve a wider discussion inclusive of individuals who may be otherwise inaccessible to academic researchers who use more traditional or in-person community engagement strategies.

The hashtags *#AcademicTikTok*, *#LearnOnTikTok*, and *#ProfessorsofTikTok* are just a few of the many examples demonstrating knowledge dissemination and educational content creation already occurring on the platform. These hashtags link researchers, educators, health care providers, and health care users together with diverse audiences around the world, creating innovative digital knowledge exchange. Some of the content uses humor to critically reflect on academia—for example, poking fun at the peer-review publication process—whereas other videos serve to share the results of a research paper and educate or even recruit study participants. For example, the popular science communicator Hank Green (@hankgreen1) boasts approximately 4.5 million followers. In the context of the digital trans TikTok community, the gender-affirming surgeon Dr Sidhbah Gallagher (@gendersurgeon), who shares information about gender-affirming surgeries, has approximately 165,000 followers. Peer-to-peer knowledge sharing is another form of digital knowledge exchange that often occurs on TikTok (and on other social media). Examples of peer-to-peer knowledge sharing happening on TikTok include instances of trans people sharing with other trans peers Dr Gallagher's educational videos about gender-affirming surgery or tagging their peers in the comment section to notify them about the video. In another example of peer-to-peer sharing, the trans-identified physician Dr AJ Eckert (@drajeckert) creates tiktoks containing medical information about gender-affirming hormonal therapy intended for audiences of trans health care users and other health care providers.

Over time, knowledge mobilization has shifted from a positivist, linear, top-down translation of research findings to increasingly complex, iterative approaches that include knowledge coproduction in the context of its use [42]. Said differently, the former model is being replaced by knowledge activities that endorse research as a creative enterprise defined by human relationships, the salience of quality partnerships with health care users, and power-sharing techniques [43]. In fact, knowledge mobilization strategies have more reach and better engagement when involving collaboration or co-design with stakeholders, such as patients, thereby capitalizing on already established web-based community networks [44]. Accordingly, boundaries are increasingly porous between researchers and communities under study, as evidenced by the knowledge exchange happening on TikTok. The platform engenders not only the sharing of research findings by researchers but also collaborative knowledge exchange between all TikTok users through features such as the stitch and the duet. Such collaborative approaches harness co-design between researchers and knowledge users, in turn, minimizing power differentials while engaging underrepresented voices and valuing the legitimacy of lay, or community-level, knowledge [42]. In the following section, we further discuss why trans communities specifically present unique opportunities for collaborative and community-engaged knowledge mobilization via TikTok.

Community-Engaged Knowledge Mobilization and Trans Peer-to-Peer Information Sharing

Trans people comprise approximately 0.5% to 1.3% of the general population [45]. Trans populations also confront significant barriers to health care, such as gender-affirming medicine, and experience discrimination within these settings [46]. Despite more public visibility of trans populations, trans-related stigma compounds social and health injustices [45]. Trans-related stigma occurs when ideologies, policies, and structural systems enforce a male/female binary that disadvantages trans people [47]. Trans-related stigma can occur at the systemic level through discriminatory policies and can be enacted at the interpersonal level through discrimination within individual interactions [48]. To demonstrate, the US Trans Survey found that 25% of trans-identified respondents encountered a problem with their health insurance provider denying claims related to gender-affirming medicine, with 33% experiencing at least one negative interaction with a health care provider such as verbal harassment or having to teach their own provider about appropriate care for trans people [20].

Disrupting trans-related stigma as well as sharing up-to-date empirical trans health knowledge about topics such as gender-affirming medicine are fundamental to improving health care access, health outcomes, and social conditions for this population. At the same time, social media platforms such as TikTok enable trans people to create alternative and trans-affirming knowledge that can counter stigmatizing beliefs. TikTok also provides a platform for health care providers and researchers who specialize in gender-affirming medicine to educate other clinicians and policy makers. Reflecting contemporary collaborative knowledge coproduction [42], trans communities are digitally well-connected and are active in peer-to-peer information sharing, support, and intracommunity education on the web [22,49,50]. This creates optimal conditions for community-engaged research and digital knowledge exchange between researchers, health care providers, and trans health care users. TikTok is especially popular with trans people, and *#TransTikTok #TransTok #NonbinaryTikTok and #GirlsLikeUs* are just a selection of hashtags that collate content and make peer-to-peer knowledge exchange possible. Then, through the power of the TikTok algorithm and individually curated *For You* pages, information relevant to this population is widely digitally disseminated via hashtags, duets, stitches, and the sharing and tagging of these videos between users.

However, it is important to stress the reality that trans people may seek out social media as a means of escaping a discriminatory world to connect with, and provide mutual support to, others who share a similar understanding of gender identity and expression. A 2021 scoping review found that social media and web-based communities were sources of both

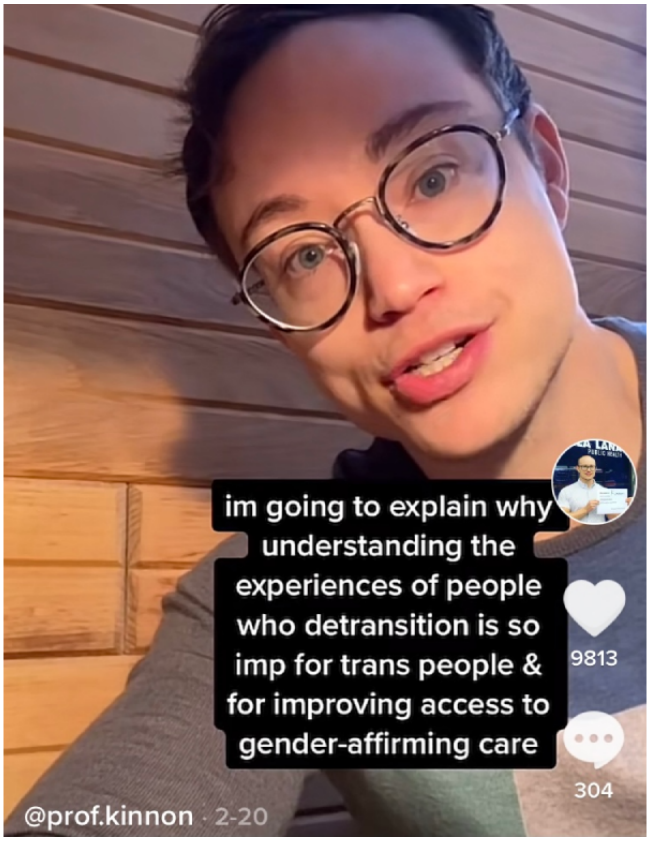
peer-based support and resilience for trans people, which in turn could buffer poor mental health outcomes experienced by this population [49]. Another 2021 scoping review identified *advocacy and education* as prominent themes of trans peer-to-peer information sharing and support [22]. Lupton [31] specifically recommends examining how TikTok and other social media platforms are used for health-related content creation and peer-to-peer sharing among young people. It is important to point out that peer-to-peer information sharing on trans health–related topics may also emerge in response to gaps in health professionals' education and competency in this domain; thus, tiktoks could be shown to clinician learners in health professions education and training. This may explain, in part, why videos concerning gender-affirming medicine generate significant attention and knowledge exchange on TikTok. In the following section, the tiktoks created by KRM (@prof.kinnon; ~13,800 followers) are shared, and analytics data are presented.

TikToking About Gender-Affirming Medicine: Digitally Mobilizing Knowledge

We sampled 13 different tiktoks created by KRM and posted them between February 19, 2021, and March 4, 2021 (Multimedia Appendix 1). These tiktoks focus on gender-affirming medicine and primarily feature his own research and related studies by other trans health researchers. Each of these videos is <52 seconds in duration and covers academic research on topics of interest to health care providers and trans people seeking and receiving hormones and surgeries. Figure 1 shows one of these tiktoks about gender-affirming medicine eligibility assessments [51] that, to date, has received >70,000 views and 8746 positive endorsements (eg, likes or hearts). For our purposes, we simply present the number of views and shares for this compiled group of 13 tiktoks. At the time of writing (May 7, 2021), these videos were cumulatively viewed >378,301 times and shared >1313 times. Taken together, the creation of these short videos alongside much engagement from web-based trans communities via *#transtiktok* demonstrates the potential for innovative, digital knowledge mobilization and exchange on TikTok.

TikTok analytics data are included within each video, updated daily, and provided to every TikTok user with >1000 followers, enabling opportunities for researchers to monitor and measure the impact of their videos. These data provide researchers with important analytics metrics to track the impact and reach of their videos, such as the average view time, the total number of likes or hearts, and very basic viewer demographics (gender and country). Although counterintuitive, it is not uncommon for videos posted at an earlier date to gain traction with the algorithm at a later date, be pushed to users' *For You* page, and subsequently go *viral*—a term used to denote a trending video. In such cases, a video has been liked, commented on, and/or shared by thousands of TikTok users, and the algorithm is sensitive to this activity, thus prompting the algorithm to showcase it on other users' *For You* page [38]. The more that users watch and engage with individual tiktoks, the more the algorithm shows the video to other users with a history of engaging with videos of similar content. Thus, videos can gain TikTok users' attention at discrete periods unrelated to the original time and date posted.

Figure 1. "Regret" and gender-affirming care.



Considerations for Researchers Planning to Use TikTok for Digital Knowledge Mobilization and Exchange

Overview

Although researchers in past years primarily used mass media, including newspapers or radio, to communicate the latest scientific research study findings, the public's newfound reliance on social media to access information about current events has increased. A survey commissioned by the Canadian Journalism Foundation found that 52% of Canadians and 48% of Americans aged ≥18 years use social media for news-related information [52]. In another example, survey data collected by Pew Research suggest that approximately 1 in 5 American adults access political news from social media, with 48% of those aged 18 to 29 years reporting social media as their main news source [53]. The demographics of these individuals who use social media as the main news source also tend to be "younger, are less likely to be white and have lower levels of education" [53]. Within the new digital age of disinformation and *fake news*, social media engagement is a contentious subject, particularly for academic researchers [54]. Social media platforms have clearly become indispensable tools for disseminating empirical knowledge widely and equitably.

TikTok is an innovative digital platform embedded with possibilities for transformative, collaborative knowledge mobilization and community-engaged knowledge exchange between researchers, underserved health care users, and their health care providers. As we have shown in discussing trans populations as an exemplar, TikTok presents unique opportunities for collaborative and community-engaged knowledge mobilization, and it may effectively promote critical dialog and information sharing to advance health care access for populations who experience poor health outcomes and barriers to care. However, there are several issues that researchers need to consider before launching their first tiktoks into the digital world. These can be summarized as follows: (1) TikTok analytic tools to measure knowledge mobilization and exchange, (2) population-specific safety concerns on social media, (3) the spread of disinformation and the need for legislation and regulation, (4) reaching those who have barriers to internet access, and (5) commercialization and intellectual property concerns.

Of relevance to those interested in using TikTok analytics to measure and evaluate the reach and impact of their videos, Basch et al [40] pointed out the challenge in distinguishing between the number of video views from viewers. Although TikTok provides users with viewer analytics data, it is not possible to identify when individual viewers watch a video multiple times or whether the videos are watched to completion. TikTok analytics do track the overall average watch time of each video; however, disaggregating these analytics data is not currently possible. These are platform-specific confounding factors that researchers must be aware of when seeking to measure and evaluate their digital knowledge mobilization activities on TikTok.

When leveraging TikTok for knowledge mobilization and exchange, there may be population-specific psychosocial developmental or safety considerations depending on the target stakeholder or equity-seeking group. For instance, youth aged <25 years comprise most of the TikTok users; however, social media use is positively correlated with depression, anxiety, and sleep disturbances in young people [55]. Some research cautions that youth engaging with health-related content on social media tend to compare their bodies with popular social media influencers [31] and that there are psychosocial risks, given the publicized evaluation metrics of likes, shares, and comment sections that allow users to openly criticize the video creator [34]. Similarly, young trans people use TikTok, and trans people of all ages regularly experience trans-related stigma on the web and in person. Trans people further report high rates of body image concerns and eating disorders, and a literature review published in 2016 concluded that body dissatisfaction is central to trans people's psychosocial distress [56]. Moreover, Maly [39] noted that it is through the algorithm that these metrics contribute to the creation of a populist voice whereby followers, likes or hearts, and shares become *political facts*. However, a study of a Facebook group suicide prevention program with youth concluded that young people could be safely and effectively engaged on social media, even on psychologically sensitive topics such as suicide [57].

A possible solution to account for these population-specific safety risks is through the piloting of digital knowledge mobilization products, or knowledge exchanges, on other social media that offer private and anonymous engagement options. These alternative options may include private Facebook groups, Twitter, or Discord. Given TikTok's powerful international algorithm and its stitch and duet video editing features, this platform may not always be the most accessible option for marginalized communities who are not public about their identities or who confront harassment on the web, such as trans people.

Although TikTok presents opportunities for collaborative knowledge exchange in the spirit of health equity and social justice, the app itself is not impervious to wider social discrimination. TikTok has admitted to *shadow-banning* or otherwise suppressing content created by people from minoritized groups, such as lesbian, gay, bisexual, trans, intersex, fat, disabled, and racialized people [58,59]. According to TikTok, creators from these groups may be more vulnerable to cyberbullying; therefore, the hashtags often used within these communities are rendered less discoverable by the algorithm [59]. In the wrong hands, TikTok's stitch and duet film editing features could be used maliciously to misappropriate the original video's message. Furthermore, the possibility of facing swapping or creating *deep-fake* videos has been noted [38]. Despite TikTok's description of the functions of shadow-banning content, this measure may result in restrictions on the reach and impact of knowledge mobilization and exchange. Furthermore, there also exists disinformation and hateful content on the app, including expressions of transphobia [60]. Further analyses on the use of TikTok to share hate and right-wing propaganda has been demonstrated in the study by Weimann and Masri [60]. Minoritized populations have been misrepresented, and their

experiences misinterpreted in knowledge creation and dissemination long before TikTok and other social media. This is a harsh reality that all researchers working with marginalized communities must be accustomed to thinking about and developing plans to account for these risks.

An obvious approach to disrupt antiscience, disinformation, or other related ideologies harmful to trans and equity-seeking populations is for academic researchers, health care providers, and educators to debunk these beliefs with research [54]. This could be done using TikTok's stitch or duet video editing options to directly respond to disinformation with a new TikTok video or by downloading the original video and engaging critically with its content in the context of teaching and learning. Health professions educators should note that tiktoks can be shown to learners by linking the TikTok's URL address or by downloading the video to a device such as a smartphone. Once stored on a smartphone, a TikTok video can be uploaded and shared on other social media platforms. Of note, each of these options risks amplifying harmful disinformation as the algorithm is sensitive to such activities; in response, the original TikTok video could be pushed to be viewed by more users of the TikTok app itself. Better legislation and regulation of TikTok are clearly required to prevent the spread of disinformation. In cases where researchers are aware that the content of their video may ignite significant controversy, it is possible to disable the duet and stitch options, the option for others to download the video, and turn off the comment section, thereby foreclosing the potential for tampering with the original, intended messages. Cautious researchers may consider disabling these features permanently, although in doing so, some of the collaborative knowledge co-design opportunities with others via stitch and duet would be limited.

Like other economically marginalized populations, trans people have varying access to smartphones and reliable internet connections because of individual socioeconomic and regional factors. Given the existing literature on the digital divide [61], it is plausible that trans adults living in rural regions with sparse access to the internet, those with limited resources to purchase a smartphone, and those who are older may face barriers to participating in these forms of digital knowledge exchange occurring on TikTok. The US Trans Survey of 27,715 respondents indicated that 15% of the sample were aged 45 to 65 years, with 3% aged >65 years [20]. Moreover, 30% of the participants who responded to an income question reported an annual household income <US $20,000 per year, with half of those reporting an annual household income < US $10,000 per year. Furthermore, 3% of the sample who responded to a technology question reported not having a cell phone, which accounted for 882 people [20]. Given that the US Trans Survey was administered in a fully web-based format, this prevalence may be even higher.

Some researchers may be hesitant to share study results using a commercial, for-profit app or may hold reservations because of intellectual property concerns. To this, we note that the TikTok platform includes an intellectual property and copyright policy which "protects original works of authorship (eg, music, videos, etc.)" and it pertains broadly to the expression of users' ideas and the ways that videos or music are created rather than the ideas or facts contained within the videos or music. According to the TikTok policy, using copyrighted content may violate TikTok's policies and can be removed [62]. Diligent researchers may consider consulting in-depth, professional intellectual property legal advice, as well as their own institutional affiliations and funding bodies, before communicating their study findings on TikTok. At the same time, community-engaged researchers who are planning to share study results on TikTok may want to proactively discuss with and elicit feedback from community members and/or health care providers at the early stages of their projects (before dissemination phases). In fact, TikTok could be used as a digital community engagement tool to solicit research topic ideas from health care users, and other stakeholders, before commencing a community-engaged research study.

Despite this myriad of considerations, TikTok may still be a potential digital mechanism for researchers to share knowledge with international audiences or to target specific knowledge user groups, such as health care providers or underserved health care users. We stress that, given the shifting landscape of social media as a primary information source, TikTok is a novel technology for collaborative digital knowledge mobilization and exchange with communities that are organized on the web and who engage in extensive peer-to-peer knowledge sharing, such as trans communities. Leveraging TikTok for community-engaged, digital knowledge mobilization and transformation is, therefore, an important strategy to achieve health equity for populations in need of better health care and social justice overall.

Acknowledgments

This was an unfunded project.

Authors' Contributions

KRM wrote and created each of the tiktoks described in this manuscript, collected analytics data, and drafted the original manuscript. Coauthors (HK and ALD) contributed equally to the manuscript's conception and contributed to cowriting and editing the final version submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
TikToks on trans health.
[DOCX File , 14 KB - jmir_v23i12e30315_app1.docx]

References

1. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021 Mar;3(3):175-194. [doi: 10.1016/s2589-7500(20)30315-0]
2. Littman L. Rapid-onset gender dysphoria in adolescents and young adults: a study of parental reports. PLoS ONE 2018;13(8):e0202330 [FREE Full text]
3. Restar A. Terrific work from the legendary @ButNotTheCity. Twitter. 2020. URL: https://twitter.com/Dr_ArjeeRestar/status/1292871430812717059 [accessed 2021-04-20]
4. Restar A. The function and beauty of being a faculty is to speak truth and SCIENCE to power. Twitter. 2019. URL: https://twitter.com/Dr_ArjeeRestar/status/1164513692127584268 [accessed 2021-02-12]
5. Ashley F. 'Rapid-Onset Gender Dysphoria'. Twitter. 2020. URL: https://twitter.com/ButNotTheCity/status/1292817626868187137 [accessed 2021-04-20]
6. Restar AJ. Methodological critique of Littman's (2018) parental-respondents accounts of "Rapid-Onset Gender Dysphoria". Arch Sex Behav 2020 Jan 22;49(1):61-66 [FREE Full text] [doi: 10.1007/s10508-019-1453-2] [Medline: 31011991]
7. Ashley F. A critical commentary on 'rapid-onset gender dysphoria'. Sociol Rev 2020 Aug 10;68(4):779-799. [doi: 10.1177/0038026120934693]
8. Literat I. "Teachers Act Like We're Robots": TikTok as a window into youth experiences of online learning during COVID-19. AERA Open 2021 Feb 25;7:233285842199553. [doi: 10.1177/2332858421995537]
9. Fiesler C, Beard N, Keegan B. No robots, spiders, or scrapers: legal and ethical regulation of data collection methods in social media terms of service. Proc Int AAAI Conf Web Soc Media 2020;14(1):187-196 [FREE Full text]
10. Noble S. Algorithms of Oppression: How Search Engines Reinforce Racism. New York: NYU Press; 2018.
11. Pasquale F. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press; 2015.
12. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. J Acquir Immune Defic Syndr 2016 Aug 15;72 Suppl 3:235-242 [FREE Full text] [doi: 10.1097/QAI.0000000000001088] [Medline: 27429189]
13. Bauer GR, Scheim AI, Pyne J, Travers R, Hammond R. Intervenable factors associated with suicide risk in transgender persons: a respondent driven sampling study in Ontario, Canada. BMC Public Health 2015 Jun 02;15(1):525-515 [FREE Full text] [doi: 10.1186/s12889-015-1867-2] [Medline: 26032733]
14. Colizzi M, Costa R. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. Neuropsychiatr Dis Treat 2016 Aug;Volume 12:1953-1966. [doi: 10.2147/ndt.s95310]
15. Scheim AI, Perez-Brumer AG, Bauer GR. Gender-concordant identity documents and mental health among transgender adults in the USA: a cross-sectional study. Lancet Pub Health 2020 Apr;5(4):196-203. [doi: 10.1016/s2468-2667(20)30032-3]
16. Greensmith H, Moore M. Health care for trans youth is under attack in UK — and it's impacting the US. Truthout. 2021. URL: https://truthout.org/articles/health-care-for-trans-youth-is-under-attack-in-uk-and-its-impacting-the-us/ [accessed 2021-12-01]
17. Kidd KM, Sequeira GM, Paglisotti T, Katz-Wise SL, Kazmerski TM, Hillier A, et al. "This Could Mean Death for My Child": parent perspectives on laws banning gender-affirming care for transgender adolescents. J Adolesc Health 2021 Jun 13;68(6):1082-1088. [doi: 10.1016/j.jadohealth.2020.09.010] [Medline: 33067153]
18. de Vries E, Kathard H, Müller A. Debate: why should gender-affirming health care be included in health science curricula? BMC Med Educ 2020 Feb 14;20(1):1-10 [FREE Full text] [doi: 10.1186/s12909-020-1963-6] [Medline: 32059721]
19. MacKinnon KR, Ross LE, Gualdron DR, Ng SL. Teaching health professionals how to tailor gender-affirming medicine protocols: a design thinking project. Perspect Med Educ 2020 Oct 16;9(5):324-328 [FREE Full text] [doi: 10.1007/s40037-020-00581-5] [Medline: 32301051]
20. James S, Herman J, Rankin S, Keisling M, Mottet L, Anafi M. The report of the 2015 U.S. transgender survey. National Center for Transgender Equality. 2016 Dec 01. URL: https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf [accessed 2021-05-01]
21. Jaffee KD, Shires DA, Stroumsa D. Discrimination and delayed health care among transgender women and men: implications for improving medical education and health care delivery. Med Care 2016 Nov;54(11):1010-1016. [doi: 10.1097/MLR.0000000000000583] [Medline: 27314263]
22. Harner V. Trans intracommunity support & knowledge sharing in the United States & Canada: A scoping literature review. Health Soc Care Community 2021 Nov;29(6):1715-1728. [doi: 10.1111/hsc.13276] [Medline: 33438797]
23. Balls-Berry JE, Acosta-Pérez E. The use of community engaged research principles to improve health: community academic partnerships for research. P R Health Sci J 2017 Jun;36(2):84-85 [FREE Full text] [Medline: 28622404]
24. Guidelines for effective knowledge mobilization. Social Sciences and Humanities Research Council. 2021. URL: https://www.sshrc-crsh.gc.ca/funding-financement/policies-politiques/knowledge_mobilisation-mobilisation_des_connaissances-eng.aspx [accessed 2021-05-07]

25. Puljak L. Using social media for knowledge translation, promotion of evidence-based medicine and high-quality information on health. J Evid Based Med 2016 Feb 28;9(1):4-7. [doi: 10.1111/jebm.12175] [Medline: 26372327]
26. Chan TM, Dzara K, Dimeo SP, Bhalerao A, Maggio LA. Social media in knowledge translation and education for physicians and trainees: a scoping review. Perspect Med Educ 2020 Feb 13;9(1):20-30 [FREE Full text] [doi: 10.1007/s40037-019-00542-7] [Medline: 31834598]
27. Gottlieb M, Dyer S. Information and disinformation: social media in the COVID-19 crisis. Acad Emerg Med 2020 Jul 24;27(7):640-641 [FREE Full text] [doi: 10.1111/acem.14036] [Medline: 32474977]
28. Uldam J, Askanius T. Online civic cultures: debating climate change activism on YouTube. I J Commun 2013;7 [FREE Full text]
29. Basch CH, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
30. Mohamad S. Creative production of 'COVID-19 Social Distancing' narratives on social media. Tijdschr Econ Soc Geogr 2020 Jun 03 [FREE Full text] [doi: 10.1111/tesg.12430] [Medline: 32836482]
31. Lupton D. Young people's use of digital health technologies in the global north: narrative review. J Med Internet Res 2021 Jan 11;23(1):e18286 [FREE Full text] [doi: 10.2196/18286] [Medline: 33427684]
32. Raun T. Video blogging as a vehicle of transformation: exploring the intersection between trans identity and information technology. Int J Cult Stud 2014 Jan 16;18(3):365-378. [doi: 10.1177/1367877913513696]
33. Jackson SJ, Bailey M, Foucault Welles B. #GirlsLikeUs: trans advocacy and community building online. New Media Society 2017 Jun 09;20(5):1868-1888. [doi: 10.1177/1461444817709276]
34. Abidin C. Mapping internet celebrity on tiktok: Exploring attention economies and visibility labours. Cult Sci J 2021 Jan 19;12(1):77-103 [FREE Full text]
35. Our mission. TikTok. 2021. URL: https://www.tiktok.com/about?lang=en [accessed 2021-02-22]
36. Sehl K. Everything brands need to know about TikTok in 2020. Hootsuite. 2020 Mar 02. URL: https://blog.hootsuite.com/what-is-tiktok/ [accessed 2021-05-07]
37. Flagship report 2020. GlobalWebIndex. 2021. URL: https://www.globalwebindex.com [accessed 2021-05-21]
38. Anderson K. Getting acquainted with social networks and apps: it is time to talk about TikTok. Library Hi Tech News 2020 Feb 08;37(4):7-12. [doi: 10.1108/lhtn-01-2020-0001]
39. Maly I. Algorithmic populism and algorithmic activism. Diggit Magazine. 2019 Nov 27. URL: https://www.diggitmagazine.com/wiki/algorithmic-populism [accessed 2021-05-21]
40. Basch CH, Fera J, Pierce I, Basch CE. Promoting mask use on TikTok: descriptive, cross-sectional study. JMIR Public Health Surveill 2021 Feb 12;7(2):e26392 [FREE Full text] [doi: 10.2196/26392] [Medline: 33523823]
41. Eghtesadi M, Florea A. Facebook, Instagram, Reddit and TikTok: a proposal for health authorities to integrate popular social media platforms in contingency planning amid a global pandemic outbreak. Can J Public Health 2020 Jun;111(3):389-391 [FREE Full text] [doi: 10.17269/s41997-020-00343-0] [Medline: 32519085]
42. Langley J, Wolstenholme D, Cooke J. 'Collective making' as knowledge mobilisation: the contribution of participatory design in the co-creation of knowledge in healthcare. BMC Health Serv Res 2018 Jul 25;18(1) [FREE Full text] [doi: 10.1186/s12913-018-3397-y] [Medline: 30045726]
43. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving Research Impact Through Co-creation in Community-Based Health Services: Literature Review and Case Study. Milbank Q 2016 Jun;94(2):392-429 [FREE Full text] [doi: 10.1111/1468-0009.12197] [Medline: 27265562]
44. Elliott SA, Dyson MP, Wilkes GV, Zimmermann GL, Chambers CT, Wittmeier KD, et al. Considerations for health researchers using social media for knowledge translation: multiple case study. J Med Internet Res 2020 Jul 23;22(7):e15121 [FREE Full text] [doi: 10.2196/15121] [Medline: 32706653]
45. Winter S, Diamond M, Green J, Karasic D, Reed T, Whittle S, et al. Transgender people: health at the margins of society. Lancet 2016 Jul;388(10042):390-400. [doi: 10.1016/s0140-6736(16)00683-8]
46. Abramovich A, de Oliveira C, Kiran T, Iwajomo T, Ross LE, Kurdyak P. Assessment of health conditions and health service use among transgender patients in Canada. JAMA Netw Open 2020 Aug 03;3(8):e2015036 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.15036] [Medline: 32857149]
47. King WM, Hughto JM, Operario D. Transgender stigma: a critical scoping review of definitions, domains, and measures used in empirical research. Soc Sci Med 2020 Feb 21;250 [FREE Full text] [doi: 10.1016/j.socscimed.2020.112867] [Medline: 32163820]
48. Hughto JM, Reisner SL, Pachankis JE. Transgender stigma and health: a critical review of stigma determinants, mechanisms, and interventions. Soc Sci Med 2015 Dec;147:222-231 [FREE Full text] [doi: 10.1016/j.socscimed.2015.11.010] [Medline: 26599625]
49. Kia H, MacKinnon K, Abramovich A, Bonato S. Peer support as a protective factor against suicide in trans populations: a scoping review. Soc Sci Med 2021 Jun;279. [doi: 10.1016/j.socscimed.2021.114026] [Medline: 34010778]
50. Andalibi N, Lacombe-Duncan A, Roosevelt L, Giniel C, Wojciechowski K. LGBTQ persons? use of online spaces to navigate conception, pregnancy, and pregnancy loss: an intersectional approach. Transact Comput-Hum Interact (TOCHI) 2021:In Press (forthcoming) [FREE Full text]

51. MacKinnon KR, Grace D, Ng SL, Sicchia SR, Ross LE. “I don’t think they thought I was ready”: how pre-transition assessments create care inequities for trans people with complex mental health in Canada. Int J Ment Health 2020 Jan 26;49(1):56-80. [doi: 10.1080/00207411.2019.1711328]
52. On behalf of the CJF, Maru/Matchbox interviewed a representative sample of 1,516 Canadians and 1,523 Americans. The Canadian Journalism Foundation. 2019. URL: https://cjf-fjc.ca/sites/default/files/CJF_Misinformation%20Study_CA-US.pdf [accessed 2021-02-21]
53. Americans who mainly get their news on social media are less engaged, less knowledgeable. Pew Research Centre. 2020 Jul 30. URL: https://www.journalism.org/2020/07/30/americans-who-mainly-get-their-news-on-social-media-are-less-engaged-less-knowledgeable/ [accessed 2021-02-02]
54. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. Nature 2021 Apr;592(7855):590-595. [doi: 10.1038/s41586-021-03344-2] [Medline: 33731933]
55. Primack BA, Escobar-Viera CG. Social media as it interfaces with psychosocial development and mental illness in transitional age youth. Child Adolesc Psychiatr Clin N Am 2017 Apr;26(2):217-233 [FREE Full text] [doi: 10.1016/j.chc.2016.12.007] [Medline: 28314452]
56. Jones BA, Haycraft E, Murjan S, Arcelus J. Body dissatisfaction and disordered eating in trans people: a systematic review of the literature. Int Rev Psychiatry 2016 Nov 30;28(1):81-94. [doi: 10.3109/09540261.2015.1089217] [Medline: 26618239]
57. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing social media-based suicide prevention messages in partnership with young people: exploratory study. JMIR Ment Health 2017 Oct 04;4(4):e40 [FREE Full text] [doi: 10.2196/mental.7847] [Medline: 28978499]
58. Rosenblatt K. Months after TikTok apologized to Black creators, many say little has changed. NBC News. 2021 Feb 09. URL: http://www.nbcnews.com/pop-culture/pop-culture-news/months-after-tiktok-apologized-black-creators-many-say-little-has-n1256726 [accessed 2021-08-24]
59. Botella E. TikTok admits it suppressed videos by disabled, queer, and fate creators. Slate. 2019 Dec 04. URL: https://slate.com/technology/2019/12/tiktok-disabled-users-videos-suppressed.html [accessed 2021-08-24]
60. Weimann G, Masri N. Research Note: spreading hate on TikTok. Studies Conflict Terrorism 2020 Jun 19:1-14. [doi: 10.1080/1057610x.2020.1780027]
61. 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 Feb;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
62. Copyright policy. TikTok. 2021. URL: https://www.tiktok.com/legal/copyright-policy?lang=en [accessed 2021-02-22]

Abbreviations

WHO: World Health Organization

Edited by R Kukafka, G Eysenbach; submitted 10.05.21; peer-reviewed by S Katz Kattari, D Chen; comments to author 30.07.21; revised version received 30.08.21; accepted 02.11.21; published 09.12.21.

Please cite as:
MacKinnon KR, Kia H, Lacombe-Duncan A
Examining TikTok’s Potential for Community-Engaged Digital Knowledge Mobilization With Equity-Seeking Groups
J Med Internet Res 2021;23(12):e30315
URL: https://www.jmir.org/2021/12/e30315
doi:10.2196/30315
PMID:34889739

Viewpoint

Transporting an Artificial Intelligence Model to Predict Emergency Cesarean Delivery: Overcoming Challenges Posed by Interfacility Variation

Joshua Guedalia[1*], MBA, PhD; Michal Lipschuetz[1,2*], RN, PhD; Sarah M Cohen[2], MPH; Yishai Sompolinsky[2], MD; Asnat Walfisch[2], MD; Eyal Sheiner[3], MD; Ruslan Sergienko[4], MA; Joshua Rosenbloom[2], MD; Ron Unger[1], PhD; Simcha Yagel[2], MD; Hila Hochler[2], MD

[1]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel

[2]Division of Obstetrics & Gynecology, Hadassah Medical Organization and Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel

[3]Department of Obstetrics and Gynecology, Soroka University Medical Center, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[4]Department of Public Health, School of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[*]these authors contributed equally

Corresponding Author:
Michal Lipschuetz, RN, PhD
Division of Obstetrics & Gynecology
Hadassah Medical Organization and Faculty of Medicine
Hebrew University of Jerusalem
Churchill Avn. 8
Jerusalem, 9765415
Israel
Phone: 972 25841111
Email: michal.lipschuetz@gmail.com

Abstract

Research using artificial intelligence (AI) in medicine is expected to significantly influence the practice of medicine and the delivery of health care in the near future. However, for successful deployment, the results must be transported across health care facilities. We present a cross-facilities application of an AI model that predicts the need for an emergency caesarean during birth. The transported model showed benefit; however, there can be challenges associated with interfacility variation in reporting practices.

(J Med Internet Res 2021;23(12):e28120) doi:10.2196/28120

KEYWORDS

machine learning; algorithm transport; health outcomes; health care facilities; artificial intelligence; AI; ML; pregnancy; birth; pediatrics; neonatal; prenatal

Introduction

The integration of artificial intelligence (AI) into health care is expected to significantly influence the practice of medicine [1-4]. Machine learning (ML) as a modeling strategy is an attractive option for characterizing and predicting complex biological phenomena [5].

Critics of AI applications note that the applications are primarily based on retrospective research, with insufficient focus devoted to "real-life" implementation and verification of reproducibility in clinical practice [5,6]. For example, an ML prediction algorithm developed in an urban tertiary care center with a diverse patient population may be unsuitable for a community hospital treating a homogenous population according to local protocols.

Therefore, transporting AI models across health care facilities is critical to effectively translating AI research into medical practice [7]. In this study, we aimed to investigate the validation of a model to predict the need for an emergency caesarean during birth, the critical challenge stemming from interfacility variation in subjective measurements, and to devise a method to address this challenge.

Methods

In brief, we developed 2 ML models to predict the risk for emergency caesarean delivery (for a detailed description of the methods and model features, see Multimedia Appendix 1 and [8]). The first model was designed to be used at admission to the labor and delivery unit (admission model); the second model was designed for use during labor, integrating additional data that accumulate as labor progresses (labor progression model). These additional data supplementing the model allow for more accurate prediction. Both models will alert the staff of the likelihood that a parturient might require an emergency caesarean delivery, allowing for the preparation of staff and patient.

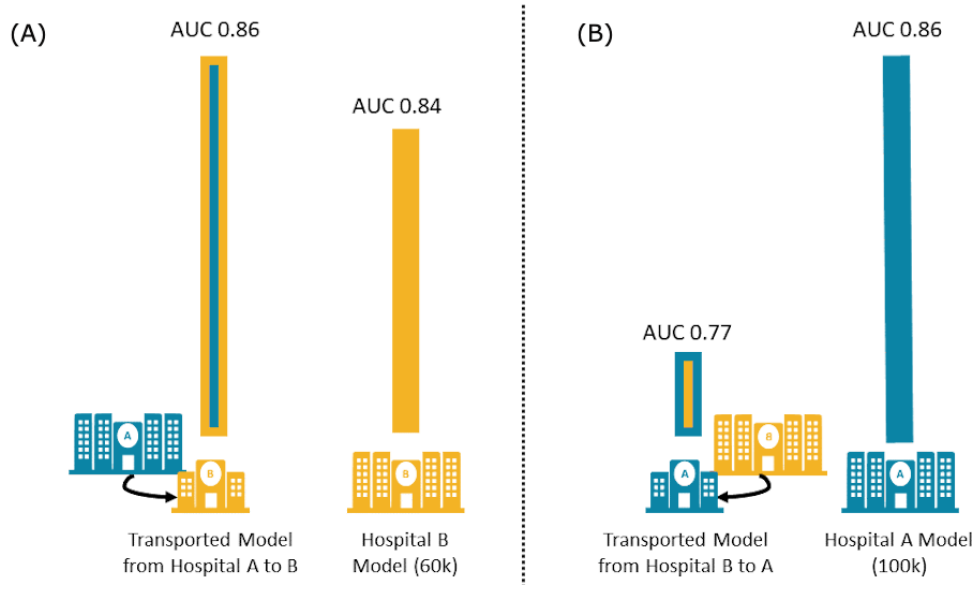
The models were trained using data from approximately 100,000 births at Hospital A. We extracted multiple data features from individual parturient electronic medical records (EMRs), totaling approximately 11 million data points. The institutional review boards at Hadassah Hebrew University Medical Center and Soroka Medical Center approved the study.

Both models were able to predict the need for emergency caesarean delivery, with the admission model achieving an area under the curve (AUC) of 0.82 and the labor progression model showing an increased performance, with an AUC of 0.86.

Having created and trained an ML-based model at a given health care facility, model transport can provide a smaller facility its benefits, without the large stored medical records or the expense and expertise required for development. However, care must be taken to monitor how the transport may affect the performance of the models, given differences in populations or settings.

We compared the prediction performance of the models trained and tested at Hospital A when transported to a second facility, Hospital B, where they were tested on data from approximately 60,000 births. Both the admission and labor progression models transported from Hospital A showed comparable prediction performance at Hospital B. Figure 1A illustrates the transport and performance of the labor progression model (see Multimedia Appendix 2 for the hospital characteristics and Multimedia Appendix 3 for the AUCs and 95% CIs of all models).

Figure 1. (A) Comparing the performance of Hospital A labor progression model (in blue) transported to Hospital B (yellow/blue bar) versus Hospital B local model (in yellow) and (B) Comparing the performance of Hospital B labor progression model (in yellow) transported to Hospital A (blue/yellow bar) versus Hospital A local model (in blue). AUC: area under the curve.



We then reversed the process and retested the success of transporting the models, by training the models at Hospital B and testing the prediction accuracy at Hospital A. Although the admission model trained at Hospital B provided similar levels of prediction at Hospital A, the labor progression model showed a reduced level of prediction (AUC 0.77 vs AUC 0.84; Figure 1B). We examined the model features to determine the cause of this decreased performance (see Multimedia Appendix 1).

Two important measurements of labor progression are fetal head station and cervical dilation. Fetal head station denotes the fetal descent within the maternal pelvis based on the position of the fetal head in centimeters above (–) or below (+) the maternal ischial spines [9]. Cervical dilation refers to the opening of the maternal uterine cervix, in centimeters, from closed cervix (0 cm) to full cervical dilation (10 cm). These 2 measurements represent the primary features of the progress of the birth; how rapidly descent and dilation progress depends on several factors, including parturient parity, medical history, pelvic anatomy, the size of the fetus, and the position of the fetus at the time of labor [10]. Results are operator-dependent, and measurements can vary between facilities based on local protocols and practice habits [11].

We identified a difference between the 2 facilities in fetal head station measurements used by the labor progression model. Specifically, we found that the dispersion and central tendency of this variable, as stratified to cervical dilation, differed between the 2 hospitals: Data from Hospital A were widely distributed across the full –3 to +3 scale, while those from Hospital B were more concentrated around –2 to +2. This difference may explain the reduced performance when transporting from Hospital B,

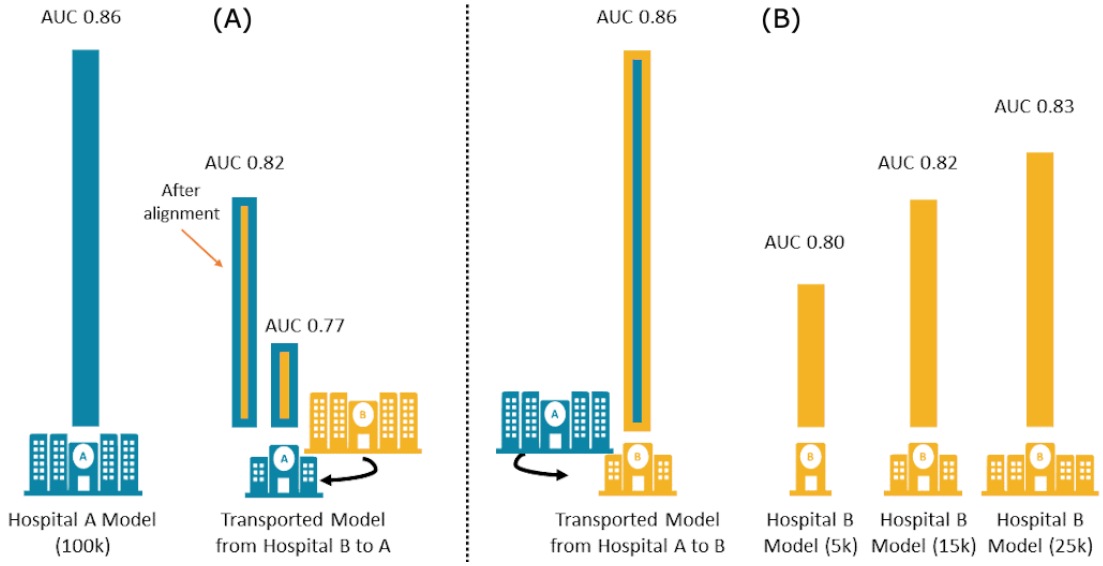
while no reduction in performance was observed when transporting from Hospital A.

In order to overcome this disparity, we evaluated the patterns of distribution of fetal head station as distributed across the dilation. We aligned the station within the distribution of the cervical dilation in order to encompass both approaches. This partly adjusted for the variation and improved the cross-facility prediction (AUC 0.82; Figure 2A; see Multimedia Appendix 1 and Multimedia Appendix 3 for the AUCs and 95% CIs of all models).

This difference highlights the difficulties introduced by discrepancies in reporting practices between facilities. Alignment can resolve some disparities, but here, it only partly recouped model performance.

To further evaluate whether our labor progression model could potentially benefit an even smaller facility, we simulated a hospital with a smaller EMR. The 100,000-case Hospital A model transported to Hospital B showed better performance (AUC 0.86) than a Hospital B model based on small samples of 5000 (AUC 0.80), 15,000 (AUC 0.82), and 25,000 (AUC 0.83) cases, emphasizing the benefit that can accrue to a smaller facility from a model trained at a larger facility and that the additional benefit decreases as the size of the available local EMR grows (Figure 2B).

Figure 2. (A) Comparing the performance of Hospital B labor progression model (in yellow) transported to Hospital A versus Hospital B model after alignment adjustments transported to Hospital A (blue/yellow bars) versus Hospital A local model (in blue) and (B) Comparing the performance of Hospital A labor progression model transported to Hospital B (yellow/blue bar) versus Hospital B local models trained on progressively larger local electronic medical record (EMR) data sets of 5000, 15,000, and 25,000 (in yellow). AUC: area under the curve.



Conclusions

In conclusion, integrating ML applications into clinical medicine will require validation and transportation between medical facilities [7,12-14]. We demonstrated that ML may be applied to clinical practice and to obstetrics in particular. A big data–driven ML algorithm can be successfully transported, and a data-poor center can benefit from work performed in a larger facility.

However, transportation requires careful investigation of specific features and consideration of variations in local populations, protocols, and reporting to calibrate the system fit [7,12]. Nevertheless, model predictions are heavily dependent on the data used in training and by the variations in recording practices and protocols operative in a given health care facility. We observed that the more detailed labor progression model, when trained without accounting for reporting differences, provided a lower AUC than the admission model. Although the progression model contained more detailed information on the progression of the labor and intrahospital showed benefit over the admission model, the benefit provided was lost when transporting the model to a different hospital: The transported model performance was inferior to that of the simpler model. Interfacility variation between health care centers may introduce unexpected effects into a prediction model. Generalizability and transportability among medical facilities necessitate overcoming biases via external validation and adapting the model to local protocols [15].

Successful translation of AI research into practice depends on transport across health care facilities. This can individualize health care, improve outcomes, and reduce complications across broader populations.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional methodology.
[DOCX File , 544 KB - jmir_v23i12e28120_app1.docx]

Multimedia Appendix 2
Demographic parameters of the two hospitals.
[DOCX File , 14 KB - jmir_v23i12e28120_app2.docx]

Multimedia Appendix 3
AUROC of the different models.
[DOCX File , 14 KB - jmir_v23i12e28120_app3.docx]

References

1. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020 Sep 24;3(1):126 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 34556791]
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
3. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
4. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
5. 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]
6. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689 [FREE Full text] [doi: 10.1136/bmj.m689] [Medline: 32213531]
7. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015 Mar;68(3):279-289 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.06.018] [Medline: 25179855]
8. Guedalia J, Lipschuetz M, Novoselsky-Persky M, Cohen SM, Rottenstreich A, Levin G, et al. Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries. Am J Obstet Gynecol 2020 Sep;223(3):437.e1-437.e15. [doi: 10.1016/j.ajog.2020.05.025] [Medline: 32434000]
9. Dupuis O, Silveira R, Zentner A, Dittmar A, Gaucherand P, Cucherat M, et al. Birth simulator: reliability of transvaginal assessment of fetal head station as defined by the American College of Obstetricians and Gynecologists classification. Am J Obstet Gynecol 2005 Mar;192(3):868-874. [doi: 10.1016/j.ajog.2004.09.028] [Medline: 15746684]
10. Evbuomwan O, Chowdhury YS. Physiology, Cervical Dilation. Treasure Island, FL: StatPearls Publishing; 2021.
11. Graseck A, Tuuli M, Roehl K, Odibo A, Macones G, Cahill A. Fetal descent in labor. Obstet Gynecol 2014 Mar;123(3):521-526. [doi: 10.1097/AOG.0000000000000131] [Medline: 24499749]
12. Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun 2020 Nov 09;11(1):5668 [FREE Full text] [doi: 10.1038/s41467-020-19551-w] [Medline: 33168827]
13. Hassanzadeh H, Nguyen A, Karimi S, Chu K. Transferability of artificial neural networks for clinical document classification across hospitals: A case study on abnormality detection from radiology reports. J Biomed Inform 2018 Sep;85:68-79 [FREE Full text] [doi: 10.1016/j.jbi.2018.07.017] [Medline: 30026067]
14. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/TKDE.2009.191]
15. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014 Aug 01;35(29):1925-1931 [FREE Full text] [doi: 10.1093/eurheartj/ehu207] [Medline: 24898551]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
EMR: electronic medical record
ML: machine learning

Edited by G Eysenbach; submitted 22.02.21; peer-reviewed by N Barda, T Kahlon, JA Benítez-Andrades; comments to author 02.07.21; revised version received 25.08.21; accepted 29.10.21; published 10.12.21.

Please cite as:
Guedalia J, Lipschuetz M, Cohen SM, Sompolinsky Y, Walfisch A, Sheiner E, Sergienko R, Rosenbloom J, Unger R, Yagel S, Hochler H
Transporting an Artificial Intelligence Model to Predict Emergency Cesarean Delivery: Overcoming Challenges Posed by Interfacility Variation
J Med Internet Res 2021;23(12):e28120
URL: https://www.jmir.org/2021/12/e28120
doi:10.2196/28120
PMID:34890352

Viewpoint

Going Viral: Researching Safely on Social Media

Kari Dee Vallury[1], BA, MHID; Barbara Baird[1], BA, GDip, DPhil; Emma Miller[1], BNurs, MPH, DPhil; Paul Ward[1], PGCert, MA, DPhil

College of Medicine and Public Health, Flinders University, Bedford Park, Australia

Corresponding Author:
Kari Dee Vallury, BA, MHID
College of Medicine and Public Health
Flinders University
Sturt Road
Bedford Park, 5042
Australia
Phone: 61 0872218200
Email: kari.vallury@flinders.edu.au

Abstract

Safety issues for researchers conducting and disseminating research on social media have been inadequately addressed in institutional policies and practice globally, despite posing significant challenges to research staff and student well-being. In the context of the COVID-19 pandemic and given the myriad of advantages that web-based platforms offer researchers over traditional recruitment, data collection, and research dissemination methods, developing a comprehensive understanding of and guidance on the safe and effective conduct of research in web-based spaces has never been more pertinent. In this paper, we share our experience of using social media to recruit participants for a study on abortion stigma in Australia, which brought into focus the personal, professional, and institutional risks associated with conducting web-based research that goes viral. The lead researcher (KV), a postgraduate student, experienced a barrage of harassment on and beyond social media. The supportive yet uncoordinated institutional response highlighted gaps in practice, guidance, and policy relating to social media research ethics, researcher safety and well-being, planning for and managing web-based and offline risk, and coordinated organizational responses to adverse events. We call for and provide suggestions to inform the development of training, guidelines, and policies that address practical and ethical aspects of using social media for research, mental and physical health and safety risks and management, and the development of coordinated and evidence-based institutional- and individual-level responses to cyberbullying and harassment. Furthermore, we argue the case for the urgent development of this comprehensive guidance around researcher safety on the web, which would help to ensure that universities have the capacity to maximize the potential of social media for research while better supporting the well-being of their staff and students.

(J Med Internet Res 2021;23(12):e29737) doi:10.2196/29737

KEYWORDS

cyber bullying; online bullying; research activities; occupational safety; research ethics; students; bullying; social media

Introduction

Social media is rapidly becoming a mainstream tool for the conduct and dissemination of research, health interventions, and evaluations [1]. Researchers and research students are increasingly expected to conduct and communicate their research on the web [2], including using a range of social media platforms to conduct and promote their work. Such spaces present new opportunities and risks for research. Rapid and potentially targeted recruitment and (perceived) anonymity provide access to historically hard-to-reach populations. At the same time, the boundaries between researchers' professional and personal identities have become increasingly blurred as images, information, and work are shared and searchable across platforms. As such, communication with and harassment of researchers on the web can move rapidly from public to private spaces, with a suite of personal and professional consequences that are in line with those of web-based bullying and trolling more broadly.

In this context of new risks and opportunities, research ethics processes, the literature, and guidelines are beginning to address the specific concerns associated with research participant safety and well-being in web-based and social media research. However, robust and constructive cross-institutional and interdisciplinary conversations and guidance addressing the management of and support for researcher safety and well-being

continue to be largely missing. In this paper, we argue that there is an urgent need for robust guidance on the use of social media for research, paying particular attention to the need for institutional and ethical frameworks and researcher training that address web-based safety and mental well-being. By outlining our extraordinary and challenging experience of *going viral*, along with the limited published experiences of other researchers, this paper calls for institutional and industry-wide practices that aim to keep researchers and their work safe in increasingly unavoidable web-based workspaces.

New Norms, New Risks

Under consistent pressure to meet research performance expectations in the context of time constraints, in the COVID-19 pandemic environment of limited travel and face-to-face engagement opportunities, and given the benefits of engaging with technological innovations to improve research processes, researchers increasingly occupy web-based networks and social media platforms for the communication and conduct of research. In this context, social media–enabled recruitment has never been more relevant. The reach, speed, affordability, flexibility, and potential for multidirectional communication and *sharing features* afforded by social media make it a favorable alternative to traditional research processes and their limitations [3-6]. In particular, social media has been found to be an effective tool in health research and promotion. Social media has been used to successfully recruit hard-to-reach populations and may be particularly "well-suited to research and practice on 'taboo' public health topics" [4], such as sexual health. This is partly because of the potential for anonymity on social media, along with the high number of young people present on these platforms [4,6-10]. Engaging research participants via social media can help to minimize research fatigue, facilitate engagement and retention of research participants, and contribute a richer data set than traditional methods can achieve on their own [5].

Along with these benefits, the limited (albeit growing) body of literature on using social media for research also describes challenges, including self-selection bias, engagement, and underrecruitment, along with a lack of control over the framing and sharing of content shared on the web [8,11,12]. Social media platforms have been described as *echo chambers*; users are constantly and progressively exposed to content aligned to their pre-existing belief systems, confirmation bias thus being a feature of social media use [13]. This allows for the specific targeting of messaging and advertisements beneficial to the conduct of science and health promotion; it also means politically charged or emotionally arousing content is most likely to spur engagement and *go viral* [13,14].

There are additional potential challenges associated with the use of social media in research. The absence of facial and social cues and gestures on the web that would otherwise be present in face-to-face interactions and the real or perceived anonymity that web-based interactions can afford increase the potential for interpersonal conflicts and escalation of arguments [15-17]. "Language truncation, the use of images and hashtags, results in inappropriate, inaccurate or mis-judged commentary in 140 characters" [18], which can affect the narrative that surrounds research shared on the web and limit the ability of researchers to control it [8]. Misinformation, misinterpretation, and misappropriation of research or research activities on the web could be described as somewhat of an inevitability, as is highlighted in the discussion of our own experience. Users' perceived anonymity and strength in numbers also means that communication and harassment among users can escalate rapidly, shifting from public to private and professional to personal web-based spaces [17,19,20]. Harassment on the web is not new; however, cultural and technological changes are likely to increase the risks of experiencing harassment and the speed at which *cyber mobs* rally, posing evolving challenges to researcher privacy, safety, and well-being.

Despite the myriad of challenges it poses, social media will be increasingly used by researchers who will become fluent in navigating and imagining its potential. Concurrently, these researchers will inevitably face evolving and fluent forms of harassment. As such, there is an onus on higher education and research industries and institutions to assume greater responsibility for the well-being of staff and students on the web, supporting and equipping them with the tools needed to safely navigate and effectively use these platforms and appropriately responding when harassment occurs.

Going Viral: Triumphs and Troubles

As part of the primary author's (KV) PhD research on abortion stigma in Australia, Facebook was used to recruit members of the Australian public to a web-based survey.

A number of professional, academic, and ethical challenges were faced by our research team during this process, which we share here in the hope that they will inform conversation and debate around the role of universities in better understanding, mitigating, and addressing researcher and student safety on the web.

Over 2 years, the authors developed a quantitative survey measuring abortion attitudes, knowledge, and perceived abortion stigma, which is the first of its kind to be developed and implemented in Australia. The survey tool was informed by extensive literature searching and qualitative and quantitative testing. It included, among others, a combination of items that endorsed and rejected stigmatizing abortion-related statements. The study received approval from the Flinders University ethics committee, including approval to omit all researcher names from the study documents.

Participants were recruited to the study using Facebook advertisements, which were targeted broadly at anyone living in Australia aged ≥16 years. Our ability to alter and retarget advertisements over time to ensure that the self-selected sample was as representative of the population as possible, the team's familiarity with using paid Facebook advertising and the relative speed at which recruitment could occur made recruitment via Facebook an appealing and logical choice. It may be relevant to consider that the survey was released during the height of the first round of the COVID-19 pandemic restrictions in Australia in April 2020 when other methods of recruitment were likely to be more challenging than usual.

In just 2 weeks of Facebook advertising, 3500 participants completed the survey. At this time, the advertisements were

retargeted to facilitate the recruitment of participants aged >40 years and male participants, underrepresented among the respondents. During the process of releasing these more targeted advertisements, the survey attracted the attention of a prominent antiabortion (prolife) lobby group who shared it with their membership via email and on their Facebook page. Within 48 hours, >5000 survey responses and close to 100 emails were received by the lead researcher (KV). At this time, the paid Facebook advertisements were halted, although the survey link remained live.

Comments undermining and debating the survey method and style, along with common antichoice sentiments around the "irresponsibility of women seeking abortion" and "abortion as murder" were noted as relevant social media posts. Emails to the research team and the university ethics committee contained *concerned queries* and *recommendations* for improvements, along with explicit hostility and requests to have the study ceased. McPherson et al [21] found that users who are the first to share a study (on social media) are likely to affect the composition of the resulting sample, reflecting the power and influence of individuals to amplify and influence messaging and information accuracy on the web. Our experience supports their finding, as the vast majority of the 5000 responses received in the days following the lobby group's sharing of the study reflected their otherwise minority (in Australia) strong antiabortion views.

Coordinated attempts by this lobby group to undermine rights or evidence-based laws, policies, or programs, such as those pertaining to abortion, contraception, and lesbian, gay, bisexual, transgender, queer, intersex, and asexual (LGBTQIA+) rights, including marriage equality, are common in Australia [22,23]. Along with the use of more formal lobbying channels, direct communication (often abusive) with staff involved in projects or organizations that the group does not agree with have been reported [24,25]. The potential for such a response to our research was likely amplified by the growing (at least in prominence) public mistrust in science more broadly. This was facilitated by social media and exemplified by apparent global shifts towards conservatism over recent years [18], along with the abortion decriminalization process that fueled antiabortion activism in South Australia at the time our research was taking place.

Although it has been difficult to formally track *shares* and *reposts* of the study, 3 days after the survey was shared by the antiabortion group, a prominent feminist politician, feminist author and public figure, and a number of women's health and women's rights organizations became aware of the study. To counter the perceived attempts by antiabortionists to sway the findings, these individuals and groups began sharing the survey in their networks. The survey was subsequently shared at least several thousand times across Twitter and Facebook and emailed to multiple women's health, women's rights (feminist), health provider, and lobby group mailing lists across a 4-day period. Much of the narrative around these *shares* sought to encourage people to complete the survey to balance the nature of responses received. However, in a number of social media posts, the survey purpose was misconstrued as being a tool for promoting an antichoice agenda, causing anger from proponents of abortion rights. Items asking participants to select their level of agreement or disagreement with statements reflecting common abortion-related stereotypes and antiabortion sentiments were construed as evidence that the survey was inherently antichoice, which further fueled this narrative. As such, hostility from both proponents and opponents of abortion rights was directed at the research team and the university.

Going viral resulted in 67,000 responses in 6 days, with a total of 70,051 responses received over the 3-week recruitment period. Ultimately, the final sample broadly represented the Australian public regarding support for or opposition to abortion accessibility and legality, with approximately 89% (9/10) supporting legal abortion always or mostly [26,27]. The survey link was made inactive after a week of *going viral*, 3 weeks after it was first published, as the responses received represented a mix of views and were deemed more than sufficient to facilitate a detailed and meaningful analysis. Within days of ending the recruitment, the antichoice lobby group *claimed victory* in their email newsletter, suggesting it was their campaign against the study that resulted in it being closed.

A month later, a freedom of information request was submitted to the university to seek documents related to the study, including documents that indicated the reasons for the survey being closed and the survey responses themselves. As the lead researcher (KV) was a student, their name and most of the information requested were redacted. Details regarding other members of the research team and the content of several personal emails between the lead researcher and her supervisors were provided; some of them were later published on the web by the antiabortion lobby group.

Despite such a successful recruitment process, our unpreparedness for the speed with which the survey would be shared on the web led to a number of challenges for the research team. For example, we were initially unprepared to manage (practically and emotionally) the hundreds of hostile emails, which appeared to be a coordinated attempt to shut down the project and were received in a span of a few days. Although the researchers' names were not in the public sphere, staying anonymous was a short-term solution, with the need to publish the work and findings, along with the freedom of information request, making disclosure inevitable.

A number of safety concerns arose, including concerns and uncertainty around the following: best practices for keeping safe on the web and preventing disclosure of personal details and location (of residence, in particular), the safety precautions that ought to be considered or implemented offline, and a lack of institutional capacity to provide such knowledge and support, the research team awareness of other strategies (with associated risks) that lobby or activist groups were likely to engage in, ways to balance the potential professional benefits of media interest with researcher and student well-being, and an understanding of risks and managing them to protect the university and individual reputations.

Phone and web-based meetings with the research team (because of the COVID-19 pandemic restrictions), on-campus phone-based mental health support, and the university media team and ethics committee were all available to support and

respond to the lead researcher's (KV) questions throughout the process of *going viral*. Although the responses received from individuals within the university were unanimously supportive, they were also ad hoc and sometimes conflicting. A coordinated response across departments, from media to ethics and student support, would have been beneficial in bolstering a sense of safety and clarity around how to respond and manage risks in relation to social media commentary, media requests and email communications, and threats.

Although it was deemed unlikely that web-based harassment would translate into offline risks of violence, a history of hostile activism and violence against abortion providers and supporters by antiabortion individuals and groups, both locally and abroad [28-30], contributed to heightened anxiety and fear throughout the experience. Recently, Glenza [31] described the antiabortion movement in the United States as *radicalized* and posing an *increased threat* [31]. Similarly, the decriminalization process and surrounding antiabortion campaign that occurred in South Australia, where the research team was located during the time of the research, heightened perceived risks. Overexposure to unpleasant social media commentaries and emails and comments on social media calling on people to *inundate* the research team with *concerned emails* resulted in the lead researcher (KV) experiencing both short- and long-term mental health consequences.

Researcher Harassment on the Web: An Anomaly?

There is a dearth of literature documenting research *going viral* and its impact on research outcomes and researcher well-being. Kosinski et al [6] described a project that, owing to web-based snowball sampling, successfully recruited 6 million participants over 4 years, with safety concerns not reported. Cuevas [19], a social scientist in the United States, described his experience of a large-scale, coordinated harassment campaign. It began in response to a comment Cuevas [19] had made on a social media post regarding the 2016 US presidential campaign, which rapidly moved into private and personal messages, threats, racist slurs, and false reviews, resulting in coordinated attempts to undermine his employment and family well-being. Cuevas [19] filed police charges, and the harassment was treated as a hate crime; however, he continued to experience harassment and threats to his job security. Cuevas [19] published about his experiences in the hope of giving a "voice to others who have been similarly harassed," stating in a media interview that he later received "emails from more than 60 professors from all over the world telling stories of their own" [32]. An Australian academic and antiracism activist, Dr Stephen Hagan, has also reported receiving hate mail and death threats in response to fake news reports about his work in advocating for the renaming of consumable products with racist connotations. Similar to that experienced by Cuevas [19], this hate campaign was fueled by right-wing political campaigns with racist dynamics [33]. Although neither of these harassment campaigns was initially in direct response to research activities, they were in response to web-based communication regarding their areas of expertise; in the Cuevas [19] case, the harassment rapidly became about his role as an academic and threatened it. As Viney [34] described, "academics have privileged knowledge that should be put to use in the community in a form of 'ethical academia'."

As such, activism and academia are often fundamentally intertwined. As social media becomes a vital stage for the performance and communication of science and research, the relevant social media posts made by academics may be necessarily considered to be part of their work.

Other researchers have reported harassing experiences in response to Facebook advertising, including in response to advertisements for LGBTQIA+ research participants [20,35]. Mitchell and Jones [20] reported cyberbullying in the form of Facebook comments, private messages, and voicemails to their research team, demonstrating the way harassment moves effortlessly from public to private spaces. Researchers working with marginalized communities or on marginalized social issues are most likely to face web-based harassment (usually not originating from the marginalized communities in which they are working). Research has also found that "harassment often arises in spaces known for their freedom, lack of censure, and experimental nature" [36]. This suggests that there is a particular risk for academics who are inherently working in *experimental* spaces (ie, conducting research) and who may be conducting research with or are members of marginalized communities themselves.

Trolling, defined as web-based behavior deliberately intended to antagonize or offend someone [37,38], is often intended as a silencing strategy, as was much of the response to our abortion stigma work. However, trolling is not the only method used to silence victims of web-based harassment and abuse. The advice offered to victims to help them cope with trolling is often to *not engage* with or *further provoke* abusers. However, such advice further silences the voices of victims and their stories and is situated within a victim-blaming narrative, whereby conducting work on the web is in itself deemed a *provocation* and harassment a normal response [36].

Among the Australian public, negative web-based experiences are common. In 2019, 14% of adults in Australia were estimated to have been the target of hate speech [39], and 67% had negative experiences on the web [40]. Studies with university students internationally report varied rates of cyberbullying, in part likely because of definitional and measurement variations; however, it is common for such studies in the United States, Canada, and Australia to find that between 20% and 40% of participants have experienced cyberbullying [41].

Cyberbullying and harassment result in social, mental, physical, financial, and academic consequences for victims, and these impacts are more commonly experienced by minority or marginalized individuals and communities [16,17,39,42]. *Secondary traumatic stress* may be of particular concern for researchers witnessing harassment of their target populations or for those who have experienced personal trauma themselves [35]. Studies that have addressed cyberbullying in universities (investigating contexts of web-based learning and web-based bullying of staff by students or colleagues) have found that it can lead to significant psychological harm (in terms of mental health, productivity, and engagement), occupational impacts (including risks to job security, satisfaction, and employment opportunities), and physical consequences (including the risk of violence) [17,19,41].

Ethical Considerations: Something Is Missing

The literature describing the ethical challenges associated with social media use in research are rooted in traditional ethical frameworks, with a focus on participant safety and protection. Ethical dilemmas regarding the appropriateness of the use of social media users' web-based data as *research data* and the automatic sharing of social media users' web-based behavior (including engagement with research-related content) with data companies are being increasingly addressed as they pertain to issues of consent, anonymity, and privacy [8,12]. Privacy and confidentiality risks to consenting research *participants* and nonparticipatory bystanders and the implications for participant *aftercare* (ie, the need for researchers to remain available to participants post data collection) have also been described [8,12,43]. Issues of inclusion and accessibility have also been raised, with the *digital divide* continuing to signal and exclude already-disadvantaged communities [5].

Ethics committees routinely request, as they must, detailed information about potential risks to participant safety and strategies to manage these risks. However, what is often neglected in ethics processes, the published literature on social media–based research, and institutional policies is researcher safety and well-being on the web. We acknowledge that this gap exists within a broader gap regarding researcher safety issues, described most frequently as relating to fieldwork and sensitive research, which are not new but remain inadequately addressed [5,44].

A Call for Guidance and Integrated Management of Researcher Safety on the Web

Health and social scientists and research students can face considerable risks and consequences associated with conducting research on politically contested or otherwise sensitive topics, which are characteristic of many areas of health research [45,46]. However, such risks, particularly their relevance in web-based settings, have been insufficiently acknowledged in the literature, policy, or practice. Researcher safety and work health and safety in research are most often defined in terms of risks of physical violence in field and laboratory work [47-49]. Cyberbullying policies and the literature focus largely on peer-to-peer or peer-to-staff (or vice versa) interactions.

There appears to be a dearth of comprehensive and integrated frameworks, training, and guidance for preparing research staff and students to implement and manage their work and safety on the web, both at the institutional and research levels [50]. There is limited evidence-based or regulatory guidance on the use of social media for research broadly [3,6,8,12]. This contributes to ambiguity around relevant ethical considerations and best practices, including how to interpret and apply existing ethics principles [51]. Guidelines published by The British Psychological Society note that exposure to distressing content, unsolicited attention or messages, or *derogatory attacks* may cause emotional distress and threaten researcher and institutional reputations as a result of web-based research [20]. However, descriptions of risks and implications of ethical considerations regarding public–private distinctions, confidentiality, and anonymity (among others) for researchers are not provided, nor is guidance on mitigating or managing risk and adverse events. The under- or overestimation of risks resulting from a lack of ethical and practical guidance for web-based research and inconsistent approval outcomes from ethics boards affects researchers' ability to conduct ethical web-based research and may discourage social media use in research, resulting in lost opportunities [5,6].

Research from North American universities has found that over half of their faculty members are unsure whether there are resources available to support them if they experience web-based bullying; however, they believe universities should be responsible for preventing and stopping web-based bullying [17]. Although Cuevas [19] reported a *uniformly supportive* response from the faculty to the harassment campaign against him, he also noted that he would have preferred a more assertive organizational response that would call out his attackers and deter future harassment campaigns rooted in the use of *collective power* against a public minority figure. Our own experience mirrors that of Cuevas [19] as responses to our experience were *uniformly* and personally supportive; however, there was no sense that a broader institutional response or positioning against the harassment was considered. This leads us to consider whether a desire to appear *objective* (and, likely, to appease diverse funders) mutes what should be confident, evidence-based, inclusive responses by academic and scientific institutions toward homophobic, antichoice, or other hate-fueled harassment of their staff and students.

Research institutions have a duty of care toward staff and students and, as such, an obligation to develop and implement strategies to protect researchers in the diversity of their modern workspaces. Although universities in Australia are legally mandated to hold policies addressing cyberbullying of staff, similar policies are not legally required for students [41]. A study found that although approximately 70% of Australian universities have policies relating to *bullying via computers*, less than half indicate *support for victims of bullying*, and only 20% provide *advice for students about bullying* [41]. An analysis of 465 policies at Canadian universities conducted in 2015 found that only one-third referenced cyber behaviors, and few addressed the prevention of web-based harassment [52]. Furthermore, such policies tend to focus on cyberbullying among peers or colleagues and often fail to address web-based safety management more broadly.

Failing to remain current with and address web-based safety concerns is not unique to universities. The *Guide for Preventing and Responding to Workplace Bullying* by Safework Australia [53] acknowledges the health and safety risks of bullying but fails to mention web-based harassment or bullying at all. However, as thought leaders and public institutions, it is questionable whether these gaps in universities—organizations that are designed to lead in knowledge generation and translation—are acceptable any longer.

Recommendations

In 2019, Russomanno et al [35] published what they described as "the first formal, safety and monitoring guidelines for researchers using social media" for recruitment, particularly of marginalized population group members. These guidelines recommend protection for both participants and researchers,

with a focus on minimizing, managing, and addressing negative comments and cyberbullying. Recommendations include assigning research team members to regularly administrate and monitor recruitment posts; posting advertisements for at most a 1-week period (at a time) to minimize researcher burnout; using inclusion *and* exclusion terms to minimize negative responses; restricting who can respond to or comment on public pages; frequent reviewing of Facebook policies around privacy, profanity, and reporting before recruitment to reduce the burden on research staff and decrease users' experiences of negative comments and bullying; screenshotting and reporting all negative interactions to internal review boards; organizing regular staff debriefs and team meetings to minimize compassion fatigue (secondary traumatic stress); and making a relevant referral to mental health services or resources for staff as needed [35]. Evidence also suggests that the use of both inclusion and exclusion terms when targeting Facebook advertisements could help to minimize the likelihood of cyberbullying toward both the study population and, presumably, web-based researchers [20,35]. It is the specifics of managing safety, such as those that we believe should be shared and understood widely across research institutions and ethics boards.

On the basis of our experiences, relevant guidance addressing researcher safety on the web could also speak to the following:

1. The need for the routine provision of evidence-based training in ethical issues in web-based research for both researchers and ethics committees; this could support increased confidence of institutional review boards and individual researchers in using social media research strategies effectively, along with encouraging the teaching of techniques to minimize the risk of exposure to potentially harmful content and responses
2. Information on and strategies addressing the blurring of private and professional boundaries on the web and changing notions of privacy, including the implications for researcher safety and security, and guidance on the responsibilities of institutions in cases where harassment occurs and may move through public and private spaces
3. Emphasis on the legal, practical, and ethical implications of working across various social media platforms
4. The need to understand, support, and strengthen the digital fluency and mental health risks and capacity of researchers to prevent, manage, and respond to potential harassment and bullying, including clear protocols for individual and institutional support and response when harassment does occur
5. Strategies for engaging with media, both in the more traditional sense of media training and in regard to responding and communicating on the web, ensuring such strategies are not centered around avoidance of social media or on a victim-blaming mentality
6. Understanding language use, inclusion and exclusion terms, and other platform-specific features that can help researchers to minimize risks associated with social media–based recruitment

Universities may also benefit from institution-wide efforts toward understanding and planning for the ways in which various departments and roles across the organization need to contribute to and work together toward coordinated and effective responses to adverse events.

There appears to be a consensus in the literature that guidance pertaining to web-based research ethics should be based on traditional ethical and well-being frameworks, partially to aid ethics bodies in their transition to assessing risks in these *new* web-based workspaces, particularly as overarching ethical concerns remain the same across the various locations of research [5,8,50,54]. However, the evolving risks, expectations around privacy, personal and professional boundaries, and ethical norms will necessarily generate new understandings and definitions of safety and require new applications and imaginations of existing ethical frameworks [50].

Instead of fearing the unknowns and risks of web-based research, the development of comprehensive guidance around web-based safety will help to ensure that universities and research groups have the capacity to maximize the potential of social media for research while better supporting the well-being of research staff and students. As such, we propose that the higher education sector, research institutions, and ethics bodies need to engage more fully with the emerging risks social media presents. When the potential benefits for the quality of research outcomes and for staff and student well-being are weighed against the risks of not better engaging with these issues, the urgency and importance of this work become clear.

Conflicts of Interest

BB is the coconvenor of the South Australian Abortion Action Coalition. KV holds a position as a Research Assistant with Children by Choice.

References

1. Hammer M. Ethical Considerations When Using Social Media for Research. Oncol Nurs Forum 2017 Jul 1;44(4):410-412 [FREE Full text] [doi: 10.1188/17.onf.410-412]
2. Fontaine G, Maheu-Cadotte M, Lavallée A, Mailhot T, Rouleau G, Bouix-Picasso J, et al. Communicating science in the digital and social media ecosystem: scoping review and typology of strategies used by health scientists. JMIR Public Health Surveill 2019 Sep 03;5(3):e14447 [FREE Full text] [doi: 10.2196/14447] [Medline: 31482854]
3. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitmen. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]

4. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
5. Lunnay B, Borlagdan J, McNaughton D, Ward P. Ethical use of social media to facilitate qualitative research. Qual Health Res 2015 Jan;25(1):99-109. [doi: 10.1177/1049732314549031] [Medline: 25212856]
6. 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]
7. Germain J, Harris J, Mackay S, Maxwell C. Why should we use online research methods? Four doctoral health student perspectivest. Qual Health Res 2018 Aug;28(10):1650-1657. [doi: 10.1177/1049732317721698] [Medline: 28745106]
8. Gelinas L, Pierce R, Winkler S, Cohen G, Lynch HF, Bierer BE. Nonexceptionalism, research risks, and social media: response to open peer commentaries on "Using social media as a research recruitment tool: ethical issues and recommendations". Am J Bioeth 2017 May;17(5):1-3. [doi: 10.1080/15265161.2017.1293755] [Medline: 28430072]
9. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb 01;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
10. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
11. Morrissey K, Kinderman P, Pontin E, Tai S, Schwannauer M. Web based health surveys: using a two step Heckman model to examine their potential for population health analysis. Soc Sci Med 2016 Aug;163:45-53. [doi: 10.1016/j.socscimed.2016.06.053] [Medline: 27394193]
12. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
13. Kim HS. Attracting views and going viral: how message features and news-sharing channels affect health news diffusion. J Commun 2015 Jun 01;65(3):512-534 [FREE Full text] [doi: 10.1111/jcom.12160] [Medline: 26441472]
14. Berger J, Milkman KL. What makes online content viral? J Mark Res 2012 Apr 01;49(2):192-205. [doi: 10.1509/jmr.10.0353]
15. Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MB, et al. Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci USA 2018 Sep 11;115(37):9216-9221 [FREE Full text] [doi: 10.1073/pnas.1804840115] [Medline: 30154168]
16. Jenaro C, Flores N, Frías CP. Systematic review of empirical studies on cyberbullying in adults: what we know and what we should investigate. Aggress Violent Behav 2018 Jan;38:113-122. [doi: 10.1016/j.avb.2017.12.003]
17. Cassidy W, Faucher C, Jackson M. The dark side of the ivory tower: cyberbullying of university faculty and teaching personnel. Alberta J Edu Res 2014;60(2):279-299 [FREE Full text] [doi: 10.11575/ajer.v60i2.55807]
18. Brabazon T, Redhead S, Chivaura R. Trump studies: the double refusal and silent majorities in theoretical times. Cult Stud Rev 2018 Oct 10;24(2):3-25. [doi: 10.5130/csr.v24i2.5628]
19. Cuevas JA. A new reality? The far right's use of cyber harassment against academics. American Association of University Professors. 2018. URL: https://www.aaup.org/article/new-reality-far-rights-use-cyberharassment-against-academics#.YaIk9dBBzIV [accessed 2021-12-02]
20. Mitchell KJ, Jones LM. Cyberbullying and bullying must be studied within a broader peer victimization framework. J Adolesc Health 2015 May;56(5):473-474. [doi: 10.1016/j.jadohealth.2015.02.005] [Medline: 25907647]
21. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol 2001 Aug;27(1):415-444. [doi: 10.1146/annurev.soc.27.1.415]
22. Maddox M. Why is the Australian Christian Lobby waging a culture war over LGBTQ issues? The Conversation. 2021. URL: https://theconversation.com/why-is-the-australian-christian-lobby-waging-a-culture-war-over-lgbtq-issues-127805 [accessed 2021-11-27]
23. Maddox M. Right-wing Christian intervention in a naïve polity. Polit Theol 2015 Apr 21;15(2):132-150. [doi: 10.1179/1462317x13z.00000000071]
24. Davey M. Safe Schools LGBTI program researchers 'receiving abusive emails' after Christian lobby campaign. The Guardian. 2016 Feb 24. URL: https://www.theguardian.com/australia-news/2016/feb/24/safe-schools-lgbti-program-researchers-receiving-abusive-emails-after-christian-lobby-campaign [accessed 2021-11-27]
25. Davey M. Override hate speech laws to allow marriage equality debate, urges Christian lobby. The Guardian. 2016 Feb 16. URL: https://www.theguardian.com/australia-news/2016/feb/16/override-hate-speech-laws-to-allow-marriage-equality-debate-urges-christian-lobby [accessed 2021-11-27]
26. Cations M, Ripper M, Dwyer J. Majority support for access to abortion care including later abortion in South Australia. Aust N Z J Public Health 2020 Oct;44(5):349-352. [doi: 10.1111/1753-6405.12997] [Medline: 32510724]
27. de Crespigny LJ, Wilkinson DJ, Douglas T, Textor M, Savulescu J. Australian attitudes to early and late abortion. Med J Aust 2010 Jul 05;193(1):9-12. [doi: 10.5694/j.1326-5377.2010.tb03732.x] [Medline: 20618106]

28. Allanson S. Murder on His Mind: The Untold Story of Australia's Abortion Clinic Murder. Melbourne, Victoria: Wilkinson Publishing; 2006.
29. Sifris R, Penovic T. Anti-abortion protest and the effectiveness of Victoria's safe access zones: an analysis. Monash Univ Law Rev 2018;44(2):317-340 [FREE Full text] [doi: 10.26180/5db80a22ec026]
30. Tozer K. Abortion clinic guard killer jailed for life. ABC The World Today. 2002. URL: https://www.abc.net.au/worldtoday/stories/s730342.htm [accessed 2021-01-29]
31. Glenza J. 'Radicalized' anti-abortion movement poses increased threat, US warned. The Guardian. URL: https://www.theguardian.com/world/2021/jul/08/anti-abortion-movement-radicalized [accessed 2021-10-09]
32. Silavent J. UNG professor exposes online harassment. Gainesville Times. 2018. URL: https://www.gainesvilletimes.com/news/ung-professor-exposes-online-harassment/ [accessed 2021-01-29]
33. Roe C. Anti-racism campaigner cops death threats over fake story. NITV. 2020. URL: https://www.sbs.com.au/nitv/article/2020/08/15/anti-racism-campaigner-cops-death-threats-over-fake-story [accessed 2020-08-09]
34. Viney H. The power of academic activism: using knowledge for community change. Australian Women's History Network. 2018. URL: http://www.auswhn.org.au/blog/power-academic-activism/ [accessed 2020-08-09]
35. Russomanno J, Patterson JG, Tree JM. Social media recruitment of marginalized, hard-to-reach populations: development of recruitment and monitoring guidelines. JMIR Public Health Surveill 2019 Dec 02;5(4):e14886 [FREE Full text] [doi: 10.2196/14886] [Medline: 31789598]
36. Lumsden K, Morgan H. Media framing of trolling and online abuse: silencing strategies, symbolic violence, and victim blaming. Feminist Media Stud 2017 Apr 27;17(6):926-940. [doi: 10.1080/14680777.2017.1316755]
37. Definition of "Troll". Merriam-Webster Dictionary. 2021. URL: https://www.merriam-webster.com/dictionary/troll [accessed 2021-01-29]
38. Definition of "Trolling". Cambridge Dictionary. 2021. URL: https://dictionary.cambridge.org/dictionary/english/trolling [accessed 2021-01-23]
39. Online hate speech: findings from Australia, New Zealand and Europe. eSafety Commissioner, Australian Government, Canberra. 2020. URL: https://www.esafety.gov.au/sites/default/files/2020-01/Hate%20speech-Report.pdf [accessed 2021-01-27]
40. Adults' negative online experiences. eSafety Commissioner, Australian Government, Canberra. 2020. URL: https://www.esafety.gov.au/research/adults-negative-online-experiences [accessed 2021-01-27]
41. Vaill Z, Campbell M, Whiteford C. Analysing the quality of Australian universities' student anti-bullying policies. High Edu Res Dev 2020 Feb 06;39(6):1262-1275. [doi: 10.1080/07294360.2020.1721440]
42. Peled Y. Cyberbullying and its influence on academic, social, and emotional development of undergraduate students. Heliyon 2019 Mar;5(3):e01393 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e01393] [Medline: 30963126]
43. Samuel G, Ahmed W, Kara H, Jessop C, Quinton S, Sanger S. Is it time to re-evaluate the ethics governance of social media research? J Empir Res Hum Res Ethics 2018 Oct;13(4):452-454. [doi: 10.1177/1556264618793773] [Medline: 30141738]
44. Morse JM. Ethics in action: ethical principles for doing qualitative health research. Qual Health Res 2007 Oct;17(8):1003-1005. [doi: 10.1177/1049732307308197] [Medline: 17928474]
45. Velardo S, Elliott S. Prioritising doctoral students' wellbeing in qualitative research. Qual Rep 2018 Feb 4;23(2):311-318. [doi: 10.46743/2160-3715/2018.3074]
46. Dickson-Swift V, James E, Kippen S, Liamputtong P. Doing sensitive research: what challenges do qualitative researchers face? Qual Res 2016 Aug 15;7(3):327-353. [doi: 10.1177/1468794107078515] [Medline: 22489612]
47. Roguski M, Tauri J. Key issues effecting field researcher safety: a reflexive commentary. New Zealand Sociol 2013;28(1):18-35 [FREE Full text]
48. Given LM, editor. Researcher safety. In: The SAGE Encyclopedia of Qualitative Research Methods. California: SAGE Publications; 2008.
49. Barr J, Welch A. Keeping nurse researchers safe: workplace health and safety issues. J Adv Nurs 2012 Jul;68(7):1538-1545. [doi: 10.1111/j.1365-2648.2012.05942.x] [Medline: 22309294]
50. The use of social media in recruitment to research: a guide for investigators and IRBs. Regulatory Foundations, Ethics, and Law Program, Harvard Catalyst. 2017 Mar. URL: https://catalyst.harvard.edu/wp-content/uploads/regulatory/Social_Media_Guidance.pdf [accessed 2021-02-11]
51. Kaye L, Hewson C, Buchanan T, Coulson N, Fullwood C, Devlin L. Ethics guidelines for internet-mediated research. The British Psychological Society. 2021. URL: https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research.pdf [accessed 2021-11-23]
52. Faucher C, Jackson M, Cassidy W. When online exchanges byte: an examination of the policy environment governing cyberbullying at the university level. Canadian Journal of Higher Education 2015 Apr 30;45(1):102-121. [doi: 10.47678/cjhe.v45i1.184215]
53. When online exchanges byte: an examination of the policy environment governing cyberbullying at the university level. Safe Work Australia, Australian Government. 2016. URL: https://www.safeworkaustralia.gov.au/system/files/documents/1702/guide-preventing-responding-workplace-bullying.pdf [accessed 2021-01-11]

54. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/ajph.2017.304249]

Abbreviations

LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual

Edited by R Kukafka; submitted 19.04.21; peer-reviewed by EB Davies, G Shakerinejad; comments to author 12.09.21; revised version received 18.10.21; accepted 29.10.21; published 13.12.21.

Please cite as:
Vallury KD, Baird B, Miller E, Ward P
Going Viral: Researching Safely on Social Media
J Med Internet Res 2021;23(12):e29737
URL: https://www.jmir.org/2021/12/e29737
doi:10.2196/29737
PMID:34898450

Viewpoint

Lessons Learned: Beta-Testing the Digital Health Checklist for Researchers Prompts a Call to Action by Behavioral Scientists

Rebecca Bartlett Ellis[1], PhD, RN; Julie Wright[2], PhD; Lisa Soederberg Miller[3], PhD; Danielle Jake-Schoffman[4], PhD; Eric B Hekler[5,6], PhD; Carly M Goldstein[7], PhD; Danielle Arigo[8], PhD; Camille Nebeker[5,6], EdD, MS

[1]School of Nursing, Indiana University, Indianapolis, IN, United States
[2]Exercise and Health Sciences, University of Massachusetts, Boston, Boston, MA, United States
[3]Department of Human Ecology, University of California, Davis, Davis, CA, United States
[4]Department of Health Education and Behavior, University of Florida, Gainesville, FL, United States
[5]Herbert Wertheim School of Public Health and Longevity Science, University of California San Diego, La Jolla, CA, United States
[6]The Design Lab, UC San Diego, La Jolla, CA, United States
[7]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, United States
[8]Department of Psychology, Rowan University, Glassboro, NJ, United States

Corresponding Author:
Camille Nebeker, EdD, MS
Herbert Wertheim School of Public Health and Longevity Science
University of California San Diego
9500 Gilman Drive
La Jolla, CA, 92093-0811
United States
Phone: 1 858 534 7786
Fax: 1 858 534 7786
Email: nebeker@eng.ucsd.edu

Abstract

Digital technologies offer unique opportunities for health research. For example, Twitter posts can support public health surveillance to identify outbreaks (eg, influenza and COVID-19), and a wearable fitness tracker can provide real-time data collection to assess the effectiveness of a behavior change intervention. With these opportunities, it is necessary to consider the potential risks and benefits to research participants when using digital tools or strategies. Researchers need to be involved in the risk assessment process, as many tools in the marketplace (eg, wellness apps, fitness sensors) are underregulated. However, there is little guidance to assist researchers and institutional review boards in their evaluation of digital tools for research purposes. To address this gap, the Digital Health Checklist for Researchers (DHC-R) was developed as a decision support tool. A participatory research approach involving a group of behavioral scientists was used to inform DHC-R development. Scientists beta-tested the checklist by retrospectively evaluating the technologies they had chosen for use in their research. This paper describes the lessons learned because of their involvement in the beta-testing process and concludes with recommendations for how the DHC-R could be useful for a variety of digital health stakeholders. Recommendations focus on future research and policy development to support research ethics, including the development of best practices to advance safe and responsible digital health research.

(J Med Internet Res 2021;23(12):e25414) doi:10.2196/25414

KEYWORDS

digital health; mHealth; research ethics; institutional review board; IRB; behavioral medicine; wearable sensors; social media; bioethics; data management; usability; privacy; access; risks and benefits; mobile phone

Introduction

Background

The increasingly familiar term *digital health* was defined by a United States National Institute of Mental Health working group in 2017, as the "blending of mobile health (mHealth) and health information technology (smartphones, wearable sensors, web-based resources, and electronic health records) with genetic, biological, social, and behavioral science to help consumers, clinicians, and researchers measure, manage, and improve health and productivity" [1]. For this commentary, we focus on

commercially available and research-grade digital health strategies used in behavioral health research and reflect on the ethical, regulatory, and social behavioral issues inherent in this work.

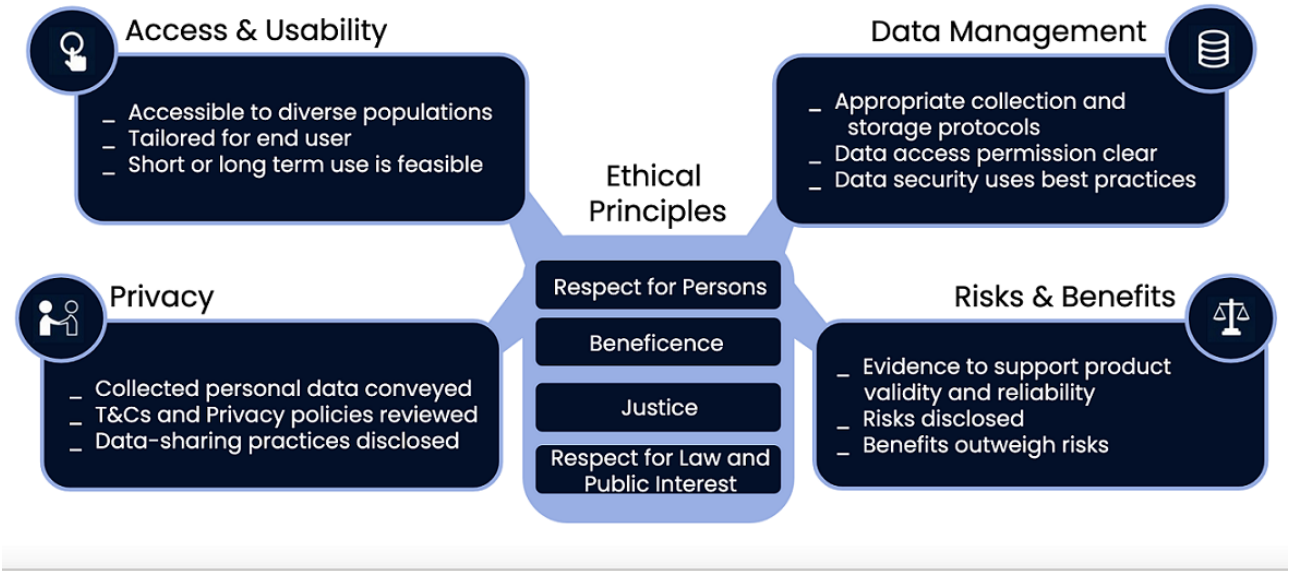
The ability to leverage digital health strategies and tools (eg, wearable and remote sensors, social media platforms, and mobile apps) to support health research holds enormous potential for behavioral and social scientists by offering accessible, scalable, and cost-effective approaches to delivering interventions to promote health behavior change, prevent disease, identify illness, and facilitate diagnosis [2]. The US National Institutes of Health began funding digital health research over a decade ago and, although the use of digital tools and strategies in research remains somewhat novel, it is escalating rapidly [3,4]. A study of United States National Institutes of Health funding of digital health research documented a 386% increase in funding of digital research between 2005 and 2015 [4]. Digital tools and strategies are increasingly used to reach populations previously understudied in biomedical research [3,5,6], including vulnerable populations with stigmatizing illnesses [7]. In the United States, the rapid onset of the pervasive technology era has preceded the development of ethical guidelines and regulatory infrastructure, which leaves researchers and potentially patients or participants vulnerable in making decisions about the selection of digital technologies and decisions about whether to participate in research. Although efforts are moving forward, these regulatory and governance gaps challenge our scientific community to inform responsible practices in digital health research, particularly new challenges and *unknown unknowns* with regards to risk assessment and data management [8,9].

Harnessing these technologies for use in health research comes with new social and ethical responsibilities to ensure that the products are effective, accessible, usable, sustainable, considerate of privacy expectations, and with good faith efforts to secure volumes of personal and sensitive health information—all of which influence the potential risks of harm and potential benefits of using such technologies [10]. The ethical responsibility to evaluate digital products for use in health research lies not only with those developing the technologies for research-grade use but also with companies that sell these products. Moreover, other key stakeholders, including researchers using digital tools and strategies to study health promotion and ethics review boards (eg, institutional review board [IRB] in the United States, research ethics board in Canada and research ethics committee [REC] in the European Union) that are charged with protecting research participants, have an important role in shaping ethical digital health research [9,11]. However, for researchers, digital health guidance for selecting tools and strategies is lacking. To promote informed decision-making in keeping with social and ethical responsibilities of those who conduct research, the Digital Health Checklist for Researchers (DHC-R) was developed [10]. This paper describes researchers' perspectives when beta-testing the first iteration of a digital health framework and checklist. Although the perspectives voiced in this paper are those of US-based researchers, we anticipate that our lessons learned, recommendations, and call to action will be relevant globally where regulations and standards to guide ethical digital health research are either limited or nonexistent.

DHC-R Background

The DHC-R is grounded in well-established ethical principles of biomedical and behavioral research and supports the following 4 domains: (1) access and usability, (2) privacy, (3) data management, and (4) risks and benefits (Figure 1) [12]. The ethical principles of respect for persons, beneficence, justice [12], and respect for law and public interest form the core of the DHC-R framework (Figure 1). The DHC-R is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License (2018-2020) and available through the Research Center for Optimal Digital Ethics (ReCODE) Health research tools webpage [13].

Figure 1. Digital health framework with examples of checklist prompts embedded within each domain (used with permission of C. Nebeker, ReCODE Health).



These 4 domains intersect with foundational ethical principles that undergird responsible practices in health research. However, researchers face a lack of consistent governance in digital health spaces. Regulations such as the US Health Information Portability and Accountability Act (HIPAA) do not always protect the granularity and volume of data and the safe storage of data derived from digital tools and strategies [14]. Moreover, not all entities conducting digital health research are governed by federal regulations created to protect research participants. This inconsistent governance creates confusion about how to assess risks of harm, manage data, and convey information to potential participants [15]. Awareness of these inconsistencies is useful when making choices about which digital solutions may work best to support the research aims and be safe for use with research participants.

Beta-Testing Approach

The DHC-R checklist was developed in 2 phases and has been reported elsewhere [10]. The DHC-R was developed in the United States and anchored to ethical principles that guide biomedical and behavioral research in the United States. Briefly, during phase 1, an expert panel was convened to review a decision-making framework published by the American Psychiatric Association, which was designed for clinicians who might prescribe or recommend a mobile app to their patients [16]. Phase 2 involved beta-testing a modified version of the American Psychiatric Association checklist for use by digital health researchers. Beta-testing involved a group of behavioral scientists who agreed to serve as the second expert panel. Expert panel members were asked to identify a recent digital health research study that they had designed and obtained IRB approval to conduct. Each panel member then applied the modified checklist, administered via the SurveyMonkey (Momentive Inc) platform, to their specific use-case study. Following completion of each domain area within the checklist (eg, privacy and data management), each panelist was asked to comment on the clarity of the prompts and whether additional criteria or content should be included (eg, *is anything missing* or *anything else to add*).

When beta-testing was completed, the lead authors (RBE and CN) reviewed all qualitative and quantitative responses and subsequently revised and published the DHC-R development process. Qualitative comments also revealed personal reflections by the panelists and a few concerns; specifically, that their prospective risk assessment may have been incomplete. With the goal of exploring their experiences during the beta-testing process, the lead authors scheduled 2 meetings with the panelists. After the first meeting, the lead authors reviewed open-ended comments and meeting notes and identified emerging themes using a thematic analysis approach [17]. During a second meeting to confirm and refine the themes, the lead authors discussed how best to share these valuable and personal insights leading to this paper's *lessons learned* summary. For context, the lessons learned presented in this paper arise from the authors' experiences working with technologies and research participants in the United States. Collectively, we conducted behavioral change intervention research that is both preventive (ie, physical activity and healthy eating) and aimed at treating chronic conditions, such as kidney disease, diabetes, and cardiovascular disease. We leverage technologies, such as wearable sensors, mobile apps, web-based social networks, and other connected technologies (ie, smart pillboxes and weight scales) in our research. Therefore, the lessons learned presented in this paper are shaped based on the United States context, including relevant United States legal protections of human subjects in research, and other relevant legislations such as US privacy laws. The goal of sharing our collective experience as panel members and behavioral scientists is to support the broader digital health research community and encourage other stakeholders to share responsibility for individually and collectively shaping ethical and responsible practices.

Results and Lessons Learned

The following 3 key themes were labeled: (1) researcher vulnerability, (2) lack of control, and (3) researchers' responsibility for human research protection. For each theme, we address what actions can be taken and future directions to consider.

Researcher Vulnerability

Overview

Reflecting on the design and implementation of their study use case left many panelists feeling vulnerable. This vulnerability stemmed from recognizing, in our role as behavioral scientists and researchers, we lacked a framework to guide the selection of the digital health tools and strategies used in our research studies. Without guidance, we needed to rely on our personal experience, technical savviness, and best judgment. Before beta-testing the DHC-R, there were no evidence-based tools or resources available to inform the selection of a digital health tools or strategy that would be appropriate for our studies and safe for participants. As researchers, we have used important and practical considerations for selecting a digital health product, such as whether the technology could be adapted to meet the needs of the research study across research sites and whether it had reasonable technical specifications. However, too often, we were not aware of important ethical aspects, and although our studies had received IRB approval, there were aspects of the digital tools and strategies we were not familiar with or knew how to consider prospectively when planning our research. Often, these issues are not known until a problem arises, which is what the DHC-R is working to preempt by helping researchers develop awareness as well as decision-making skills needed to be purposeful in advancing responsible and safe digital health research. The following response to the DHC-R beta-testing survey conveys this vulnerability:

> *Yes [I would do things differently]...That said, I also feel powerless as there is so little, I feel I understand in making decisions on use.*

Reducing Vulnerability

As researchers, we are trained to ask scientific questions and design studies to answer these questions. We have become experts in our respective disciplines; however, when we have the opportunity to use new tools or methods, we must recognize that we are novices and seek guidance from those with greater

expertise, as well as diverse experiences. We may find that working with a technologist, privacy expert, or participants can help to shed light on the *known unknowns* and potential *unknown unknowns* that are inherent when learning how to use a new research tool. Recognizing our vulnerability is necessary and humbling and will certainly lead us to become better scientists. Nevertheless, it is important to balance the fear of making the wrong decision with that of moving fast and potentially causing damage and harm. The slogan "Move purposefully and fix things" aligns with our intention of embracing vulnerability [18].

Future Directions

Our developmental research revealed that using the initial checklist can facilitate reflection on factors that influence responsible digital health research [10]. Whether this reflection process prevents one from blindly moving into work that could be risky is unknown. We also do not know if reflecting on potential risk may discourage one from pursuing important health research. Although additional research on the DHC-R is required, the checklist was useful in prompting our awareness of important considerations when selecting digital tools and strategies for health research.

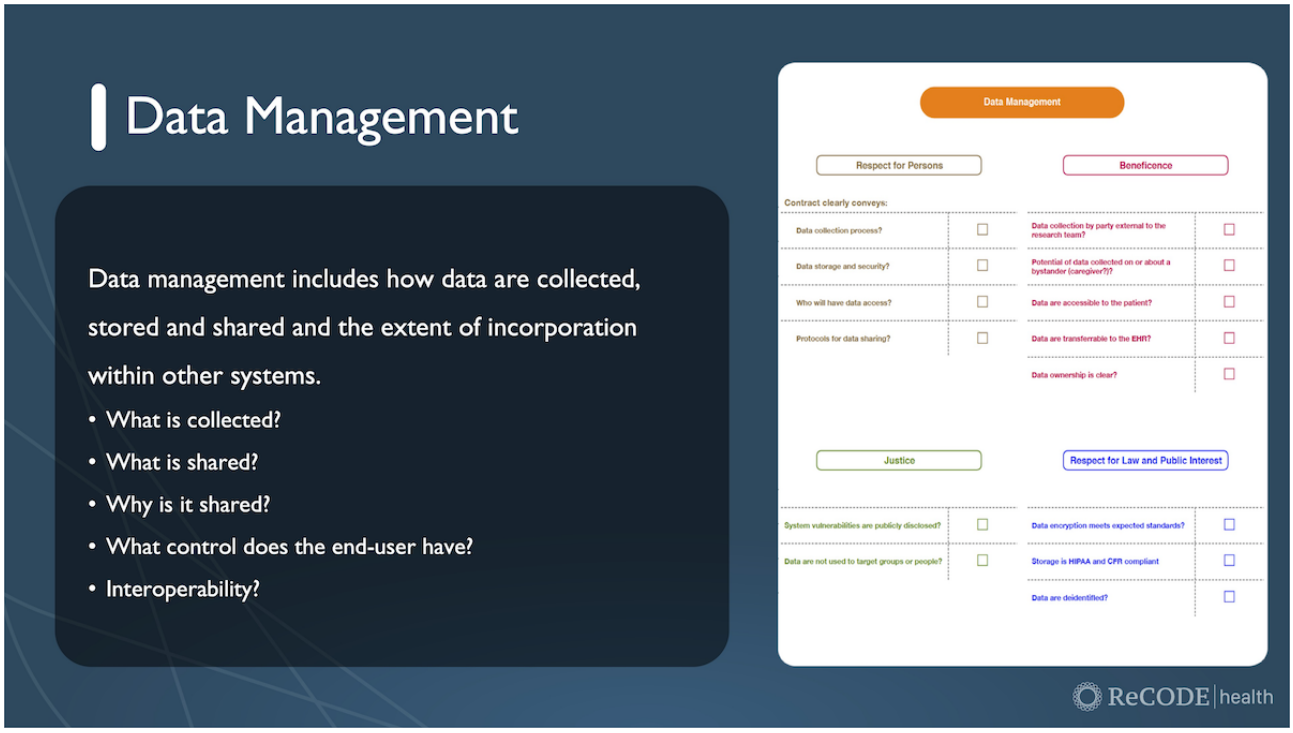
The framework (Figure 1) may reduce the vulnerability. Researchers are encouraged to consider digital tools and strategies from both the researcher and participant perspectives during the study design phase. The domain of *access and usability* prompts the consideration of the extent to which training may be required. For example, although commercial health wearables (eg, Fitbit) are designed for ease of use, participants may not intuitively know how to use real time data to make decisions about their future health behaviors. Assessing the need for and potential methods to train participants could prevent technology naivete from limiting intervention effects and influencing outcomes.

The *privacy* domain prompts researchers to take a deep dive into the vendor's terms and conditions of service to understand what data are collected, where data are stored, and how data might be shared. For example, researchers should be able to select specific variables of interest from a health technology platform (eg, location and time of day) without burdening participants to provide these data. Moreover, requiring participants to download the data file and transfer it to researchers could compromise the data quality and security. Assessing the need for a third-party data management system is also an important consideration.

Finally, items in the *risks and benefits* domain of the checklist cue researchers to think about what evidence exists to support using the technology (reliability) and the validity of the device for the particular study in question. Checklist prompts (Figure 2) guide the researcher to consider potential risks and weigh them against the potential benefits of the technology.

During beta-testing, one comment describes how future thinking may be demonstrated: "I plan to more fully consider what details to include in the informed consent, particularly about data sharing." Applying the checklist used during beta-testing was beneficial in that it prompted reflection. Moving forward, we anticipate that other researchers will find it useful to prospectively consider how a particular digital health tool or strategy aligns with their goals of doing good science and protecting research participants. For example, a researcher evaluates an app that also happens to collect GPS data, but GPS data are not needed for the research study. In this case, the DHC-R would prompt the researcher to think about either using another app, if the GPS feature cannot be *turned off*, or explaining to participants during informed consent that these data are being collected so that they can make an informed decision about whether to volunteer. Technologies on the horizon, such as the *smart toilet*, will likely continue to challenge our ability to evaluate situations in which researchers collect data that extend beyond their original research questions. The DHC-R may prove useful for decision support today, and in the future, provided it is a dynamic tool that changes with time. These choice points become particularly important if the researcher uses a third-party app that provides little or no control over the use of data outside the research study.

Figure 2. Example of data management domain from the digital health checklist (used with permission of C. Nebeker, ReCODE Health).



Lack of Control

Overview

Experiencing vulnerability stems from the awareness that researchers sometimes lack control when selecting digital technologies. For instance, researchers often sacrifice control for the benefits of using commercial products. Many commercial products make the application programming interface available to all or select approved researchers. This allows researchers to access data and deploy interventions using commercial products. For example, in a statewide dissemination of a web-based behavioral weight loss program offered through primary care clinics, participants have the option of linking a commercially available calorie counter smartphone app to their patient portal, where they are responsible for uploading weekly data on calories consumed per day, daily exercise minutes, and daily body weight [19]. Calorie counter apps are widely available, but the majority of the most popular ones do not have a researcher-accessible application programming interface. Consequently, researchers are limited in which apps they can use if they wish to access these data. At the same time, smartphone apps regularly add new (or discontinue) features, which may impact researcher access to data or make a training guide for participants obsolete when features change. In the traditional research timeline, commercial digital health products will update repeatedly during the course of a single trial, and these updates are out of the researcher's control. These updates can wreak havoc with regards to measurement and implementation and undermine the entire study design. Finally, researchers often lack control when they choose to use a commercially available digital tool or strategy because they cannot control modifications made to the data structure, algorithms, or equations, nor control whether the company stays in business, which is crucial to maintain consistent technology support and potentially data management.

Although digital tools and strategies create opportunities for researchers, the field is dynamic, making it difficult to foresee all potential risks. However, we learned that the decision-making domains of the DHC-R helped to focus panelist attention on key areas, including privacy, data management, risks and benefits, access, and usability. In this sense, the DHC-R builds awareness around important decision-making domains. The following comment made during beta-testing reflects this increased awareness:

> *There were various factors around interoperability and privacy that I did not previously consider, especially with the use of commercial apps within my study. Data from these apps weren't collected for my study, but I did request that participants try a variety of apps, which could have put them at risk for loss of confidentiality that wasn't as clear at the time of the study. Will consider this more in the future and also include transparent language in the consent forms.*

Future Directions

To move the field and researchers toward building awareness of the controllability of the digital tools and strategies, the DHC-R is best used during the planning phase of research. Not only can it be useful in prompting researchers to consider what they can and cannot control and guide development of contingency plans, the DHC-R can be used throughout the project for ongoing assessment of factors that influence responsible and ethical practices. By conducting an ongoing assessment of the products selected for a particular study, we anticipate an increased sense of control that comes with being informed. In addition to being informed, we as individuals and members of the digital health research community need to be honorable stewards of funding received to conduct behavioral health research with integrity. The responsibility to take action

resides with all in the digital health research ecosystem; however, it is not presently a norm. As such, we collectively need to commit to affecting changes in standards such that poor practices do not become the default standards. For example, if a product is not *research* friendly (eg, terms of use conflict with federal regulations for human subject protections and platform changes without notification), then those vendors become less desirable than vendors that are supportive to the research enterprise. Moreover, there is a need to prioritize the training of researchers in the pipeline such that they recognize their role in shaping policy and the dynamic development of best practices. With these added responsibilities, we recognize that research funders need to step up to support research on research ethics in the digital health sector, ideally as part of strategic planning. These collective efforts will serve to advance the development of future digital health research and raise awareness of what is acceptable verses a deal breaker.

Researcher Responsibility in Protecting Participants

Overview

Two important lessons emerged about the researcher's responsibility. The first lesson is recognizing our responsibility for due diligence in the protection of human subjects when selecting technologies and embedding them in our study design. This includes how researchers communicate and plan studies that are submitted to the IRB/REC. The second lesson is on how we take our DHC-R assessment outcome and communicate it in a way that potential participants can understand and make an informed decision about whether to participate. These 2 lessons highlight researcher responsibilities and suggest actions needed when planning future studies and how the DHC-R can assist with research design and implementation.

Due Diligence

Using the DHC-R places a shared responsibility of due diligence on both the researcher and IRB/REC—rather than relying solely on the IRB/REC or other body to determine participant safety. For example, if researchers use the DHC-R during the protocol development phase and then share that evaluation process with IRB/REC members, researchers demonstrate that they are thinking critically about designing research with prioritized participant safety. However, there were concerns that an IRB/REC may use additional knowledge about digital health tools and strategies to act more conservatively and overprotect participants, thus slowing the approval process and, possibly, stopping important health research from occurring. Moving forward, it is critical that both researchers and review boards play an equally active role in developing awareness of factors that influence study risks and benefits and learn how to use decision support tools such as the DHC-R.

We collectively agreed that researchers should not exclusively outsource the responsibility of risk assessment to the IRB/REC. Specifically, researchers should not rely on the review board as an *ethics* expert, especially in the area of rapidly changing tools and strategies used in digital health research [8]. For example, digital technology companies can be acquired by other companies, which is not often discussed in typical research consent forms. In such cases, researchers must consider how the acquisition affects research data collection, storage, and sharing while considering the long-term availability of the product for continued use. Researchers should plan to communicate this information to prospective participants during the consent process and provide updates to the currently enrolled research participants. In effect, the DHC-R helps to shape the research protocol with respect to complex aspects of data management, privacy, access, and usability, and also how to convey these concepts when crafting the information conveyed during the informed consent process.

Moving From Responsibility to Accountability

The DHC-R can help researchers identify ethical and regulatory issues that have typically been left to the review board members to identify and resolve. Moving forward, researchers as well as others who make up the digital health research sector must design an ecosystem where we collectively use checks and balances. The organization of this ecosystem is beyond the scope of this paper; however, we highlight the following steps. First, to improve informed consent, we need to make information accessible and meaningful and to present it in a way that facilitates sound decision-making. At present, researchers tend to use words to convey complex concepts that may not be useful to people with low technology and data literacy. Designing a consent process that actually informs decision-making may require that researchers partner with prospective participants and communication and design experts to learn how best to convey information. This goal is not trivial and will require dedicated funding to overhaul how we approach informed consent. For now, the DHC-R provides a framework for deep consideration of what information participants may need to make informed decisions. In addition, the DHC-R facilitates an opportunity to consider how researchers can best assist individuals in becoming better informed with respect to information contained in a terms of service agreement. Although many agree that service terms and privacy agreements are not meant to be read or understood, as researchers, we have an obligation to educate prospective participants and help them to understand what permissions are given when accepting terms to use a commercial product (eg, Fitbit and Facebook) [20]. This education can include information on privacy limitations and data management, along with the potential unknowns inherent in digital health research participation.

Future Directions

When technology is used in health research, data security and privacy responsibilities should be shared to an extent by all stakeholders, including technology developers, researchers, hosts, and participants. It is also the researcher's responsibility to inform participants when product functionality that may alter privacy and data management risks changes. Presently, institutions focus on research risk assessments to reduce legal culpability. That may be important; however, the need for institutions to help researchers to develop sound data management protocols that protect research data is critical, particularly when potentially granular personal health data collected using digital tools and strategies are not covered by United States HIPAA regulations. One beta-tester indicated the following:

> *I think it was critical that when using a commercial device, we included the creator [tech developer] as a co-investigator. This helped us think through generalizability and future applicability issues once the software and hardware are updated. I have done other studies in the past where we did *not* do this, and it was a mess. I highly recommend doing this whenever possible.*

Moving forward, we envision the DHC-R inspiring awareness of our collective responsibility in framing the ethical, legal, and policy-related decisions that support a robust digital health research culture. Ideally, researchers will be better able to identify problems and craft risk management solutions that align with developing best practices—this includes elevating awareness of products that are neither *research friendly* nor in the best interest participants who engage in our research studies.

Discussion

As researchers, we have a responsibility to (1) lead the narrative for accountable, fair, and transparent practices; (2) inform best practices by sharing experiences and conducting empirical research; and (3) help our colleagues avoid pitfalls. Following our experiences of beta-testing the checklist, which led to the DHC-R, we realized that by openly sharing our experiences, both failures and successes, we could advance the development of best practices. The DHC-R is a new decision support tool that uses a framework consisting of the following 4 domains: (1) access and usability, (2) privacy, (3) data management, and (4) risks and benefits anchored to the ethical principles of biomedical and behavioral research used in the United States [12]. Our lessons learned serve as the starting point for openly sharing successes and failures in selecting and using technologies in health research. However, with the rapidly changing digital landscape, there is an urgent need to engage researchers and the extended digital health research community in the open sharing of experiences to expand the body of lessons learned. The DHC-R provides a structure that could be useful for other responsible parties, including technology developers and IRB/REC members. Scientific and clinical progress is most likely to be achieved if the digital health research community collaborates with global stakeholders and regulatory bodies to ensure safe and innovative science.

To inform best practices, a global and multi-stakeholder perspective is needed, including traditional researchers at academic institutions, health technology companies, and citizen science initiatives. With this diverse research ecosystem comes variation in formal research training, ethics acculturation, and regulations, all of which call for elevating opportunities for engagement and dialogue [15]. Furthermore, the DHC-R was developed and beta-tested with US-based researchers; thus, the perspectives described herein may only be applicable to the United States. Engaging international stakeholders is needed to understand the extent to which the findings are applicable outside the United States.

In considering the *lessons learned*, a cluster of issues emerged surrounding the importance of, and the need for, educating stakeholders, particularly IRB/REC members and participants (who may also be patients). With that in mind, we propose *a call to action* inviting key digital health stakeholders, including review boards, participants, professional societies, and funders, to consider their respective roles in designing a responsible and ethical digital health research ecosystem.

Ethical and Regulatory Review

The IRB/RECs perform an important function in the review and approval of regulated research with human subjects. However, as technology changes and digital health studies increase, we see gaps in the ability of review boards to be well informed and make appropriate decisions [8]. The digital health community needs to collaboratively develop processes to advance ethical digital health research. Organizations that support IRB/REC professionals (eg, Public Responsibility in Medicine and Research) and recognize the unique challenges introduced by technologies in health research (eg, Society of Behavioral Medicine, American Psychological Association, American Medical Association, and Computer and Human Interaction) are well-positioned to increase educational opportunities specific to the ethical, legal, and social implications of digital health. Over the past decade, we have seen an increase in meetings that bring together experts from across sectors and disciplines, and given the increase in internet-based meetings owing to the COVID-19 pandemic, convening across professions and sectors is easier than ever. However, more work is needed to advance safe, responsible, and ethical digital health research. Researchers also have a responsibility to suggest innovative strategies for protecting patient privacy and safety to their colleagues, review boards, and professional societies. Ethical practices must be shaped as a system rather than being siloed within research areas and organizations.

Educating Participants or Patients

Educating participants about digital health research has ethical, legal, and practical considerations. From an ethical perspective, it is important to anticipate or consider how to make the informed consent process accessible, especially in studies that are completely remote. There has been some research on the process and use of e-consent [21] and efforts to improve how complex concepts are conveyed through a smartphone or tablet. We encourage efforts to involve participants (and their caregivers where appropriate) as partners and working with prospective participants to complete the DHC-R can be an approach to engage them in study planning, informing study feasibility, and assisting with risk assessment and mitigation planning.

Professional Societies

We encourage professional societies to foster a culture among members that put the ethical, legal, and social implication responsibilities of researchers at the forefront. To do so, we recommend that professional societies use the DHC-R framework and companion checklist as a decision support tool for member education efforts that encourage the conduct of socially and ethically responsible research. Furthermore, the intersection between academic researchers and industry calls for scientists to become collaborators with technologists, rather

than only being seen as end users. To accomplish this, professional societies can encourage industry partners to attend scientific meetings, not solely as vendors or industry sponsors, but as attendees. The DHC-R could be a useful framework for developing targeted education and introducing concepts of ethical decision-making in the digital health sector.

Funding and Policy

The current funding paradigm, both within the United States and abroad, typically requires compliance with regulatory mandates for human subject protection, including guidance for obtaining and documenting informed consent. However, our existing methods for obtaining informed consent do not require the assessment of the extent to which an approved consent process results in an informed participant; hence, further research is needed here. Funders can also support researchers interested in shaping digital health research policies and norms by creating a mechanism to support the empirical research that will advance evidence-based policy development. Moreover, similar to the importance of having a biostatistician on the research team, researchers should consider engaging a health technology ethicist as a coinvestigator and include aims that advance knowledge of risk assessment and meaningful consent.

Learning Communities

Learning communities that support safe and open sharing are recommended. As digital health and related new technologies continue to emerge, unknowns are expected and will continue to surface. These unknowns, when they do occur, create learning opportunities from which researchers and other stakeholders could benefit—particularly if we embrace a learning community approach. To accomplish this type of open sharing culture, a platform is required. One such platform that could serve as a model (or be the *go to* resource) is the Connected and Open Research Ethics (CORE) platform hosted by ReCODE Health [22]. To succeed, open sharing would be rewarded. For example, a novice digital health researcher who plans to use wearable sensor technologies to observe free-living daily behavior with the goal of documenting physical activity and then deploying a behavioral intervention may benefit by accessing IRB/REC-approved protocols that have been shared by a more experienced researcher. This can occur at the CORE Resource Library, where IRB/REC-approved protocols and consent documents are tagged and freely shared by and with members of the global CORE network [23,24]. Experienced researchers who share their digital health protocol to help others may be motivated through altruism or with service credits that contribute to productivity metrics for promotion. The motivation to participate as a member of the CORE should be driven by a commitment to advocate and a desire to contribute to the shaping of best practices. How that is realized will be up to our collective community efforts. Alternative metrics or rewards for researcher transparency, whereby researchers get credit for open sharing, will be how researchers are able to lead within the academic community.

Limitations

The lessons learned and call to action recommendations described in this paper are based on experiences in the United States and framed on current relevant and applicable laws, or lack thereof. Although the United States lacks a regulatory framework to guide researchers in using digital technologies in research, there are emerging state efforts that show promise. For example, in California, the recent privacy regulation will protect resident personal information collected by unregulated, for-profit businesses in cases where the United States HIPAA does not apply [25]. Internationally, there have been advances in privacy and data management. The World Health Organization has published a resolution on health technology assessments to improve decision-making and policy development [26]. In Europe, policy-driven [27,28] impact assessments have advanced from privacy impact assessments [29,30] to mandated data protection impact assessments, which are required when processing data “is likely to result in result in a high risk to the rights and freedoms of natural persons” (General Data Protection Regulation) [28,31]. The United States could learn from these efforts occurring in other countries and align with these best practices, as applicable to the United States. There is much we can learn from other countries that have been successful or are leading successful efforts in these areas.

Conclusions

Digital technologies offer potential benefits for advancing health research, yet introduce unique ethical, legal, and social implications. For researchers, due diligence in selecting technologies for research purposes requires awareness of known risks and anticipation of unknown risks to both individuals and society. Decision support tools can help researchers systematically consider the selection and use of technologies for use in health research. The DHC-R described in this commentary is one decision support tool that prompts researchers to anticipate participant privacy, consider risk of possible harms against benefits, evaluate access and usability, and review data management protocols. For researchers, the DHC-R can be used to facilitate informed research planning and preempt downstream problems associated with digital health research. Researchers who tested the DHC-R learned valuable lessons, which were synthesized to inform recommendations and a call to action directed at key stakeholders. Encouraging strategies that promote open and transparent discussions about the promise and pitfalls of digital health are required to ensure ethical and trustworthy research. Through collective action to identify issues and unmet needs within the global digital health ecosystem, an ethical and responsible digital health architecture can be realized, but only if stakeholders act now to do their part.

Conflicts of Interest

None declared.

References

1. Opportunities and challenges of developing information technologies on behavioral and social science clinical research. National Advisory Mental Health Council (NAMHC). URL: https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/opportunities-and-challenges-of-developing-information-technologies-on-behavioral-and-social-science-clinical-research.shtml#section-0 [accessed 2021-11-29]
2. Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. Ann Behav Med 2020 Apr 27. [doi: 10.1093/abm/kaaa018] [Medline: 32338719]
3. Nebeker C, Dunseath SE, Linares-Orozco R. A retrospective analysis of NIH-funded digital health research using social media platforms. Digit Health 2020 Jan 21;6:2055207619901085 [FREE Full text] [doi: 10.1177/2055207619901085] [Medline: 32030195]
4. Dunseath S, Weibel N, Bloss CS, Nebeker C. NIH support of mobile, imaging, pervasive sensing, social media and location tracking (MISST) research: laying the foundation to examine research ethics in the digital age. NPJ Digit Med 2018 Jan 15;1(1):20171 [FREE Full text] [doi: 10.1038/s41746-017-0001-5] [Medline: 31304345]
5. Armaou M, Araviaki E, Musikanski L. eHealth and mHealth Interventions for Ethnic Minority and Historically Underserved Populations in Developed Countries: an Umbrella Review. Int. Journal of Com. WB 2019 Dec 30;3(2):193-221. [doi: 10.1007/s42413-019-00055-5]
6. Nebeker C, Murray K, Holub C, Haughton J, Arredondo EM. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e87 [FREE Full text] [doi: 10.2196/mhealth.6494] [Medline: 28659258]
7. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 May;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
8. Nebeker C, Harlow J, Espinoza GR, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth 2017;8(4):266-276. [doi: 10.1080/23294515.2017.1403980] [Medline: 29125425]
9. Pagoto S, Nebeker C. How scientists can take the lead in establishing ethical practices for social media research. J Am Med Inform Assoc 2019 Apr 01;26(4):311-313 [FREE Full text] [doi: 10.1093/jamia/ocy174] [Medline: 30698793]
10. Nebeker C, Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]
11. Chevance G, Hekler EB, Efoui-Hess M, Godino J, Golaszewski N, Gualtieri L, et al. Digital health at the age of the Anthropocene. Lancet Digit Health 2020 Jun;2(6):290-291. [doi: 10.1016/s2589-7500(20)30130-8]
12. The Belmont Report - National Commission for the Protection of Human Subjects of Biomedical Behavioral Research. Department of Health, Education, and Welfare. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html [accessed 2021-11-29]
13. Research tools. ReCODE Health. URL: https://recode.health/tools/ [accessed 2021-12-13]
14. HIPAA privacy rule and public health: guidance from CDC and the U.S. Department of Health and Human Services - MMWR: Morbidity and mortality weekly report (Suppl 1). Centers for Disease Control and Prevention. 2003. URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/m2e411a1.htm [accessed 2021-11-29]
15. Nebeker C. mHealth research applied to regulated and unregulated behavioral health sciences. J Law Med Ethics 2020 Mar 28;48(Suppl):49-59. [doi: 10.1177/1073110520917029] [Medline: 32342758]
16. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 Feb 15:498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
17. Clarke V, Braun V, Hayfield N. Thematic analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. Thousand Oaks, CA: SAGE Publications; 2015:222-248.
18. 2019: a year to move purposefully and fix things -Omidyar Network. Medium. 2019. URL: https://medium.com/positive-returns/2019-a-year-to-move-purposefully-and-fix-things-c3c60747409 [accessed 2020-10-26]
19. Espel-Huynh HM, Wing RR, Goldstein CM, Thomas JG. Rationale and design for a pragmatic effectiveness-implementation trial of online behavioral obesity treatment in primary care. Contemp Clin Trials 2019 Jul;82:9-16 [FREE Full text] [doi: 10.1016/j.cct.2019.05.003] [Medline: 31063870]
20. Obar JA, Oeldorf-Hirsch A. The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. Inform Commun Society 2018 Jul 03;23(1):128-147. [doi: 10.1080/1369118x.2018.1486870]
21. Wilbanks J. Design issues in E-Consent. J Law Med Ethics 2018 Mar 27;46(1):110-118 [FREE Full text] [doi: 10.1177/1073110518766025] [Medline: 30057442]
22. ReCODE Health. URL: https://recode.health/ [accessed 2021-12-10]
23. Torous J, Nebeker C. Navigating ethics in the digital age: introducing Connected and Open Research Ethics (CORE), a tool for researchers and institutional review boards. J Med Internet Res 2017 Feb 08;19(2):e38 [FREE Full text] [doi: 10.2196/jmir.6793] [Medline: 28179216]

24. Harlow J, Weibel N, Al Kotob R, Chan V, Bloss C, Linares-Orozco R, et al. Using participatory design to inform the Connected and Open Research Ethics (CORE) commons. Sci Eng Ethics 2020 Feb 6;26(1):183-203. [doi: 10.1007/s11948-019-00086-3] [Medline: 30725245]
25. Rothstein M, Tovino S. California takes the lead on data privacy law. Hastings Cent Rep 2019 Sep;49(5):4-5 [FREE Full text] [doi: 10.1002/hast.1042] [Medline: 31581323]
26. Health technology assessment. World Health Organization. URL: https://www.who.int/teams/health-product-policy-and-standards/assistive-and-medical-technology/medical-devices/assessment [accessed 2020-12-20]
27. Magdziarczyk M. Right to be forgotten in light of regulation (EU) 2016/679 of the European Parliament and of the Council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repeali. In: Proceedings of the 6th International Scientific Conference on Social Sciences and Arts SGEM 2019. 2019 Presented at: 6th International Scientific Conference on Social Sciences and Arts SGEM 2019; April 11-14, 2019; Vienna, Austria URL: https://www.sgemsocial.org/index.php/elibrary?view=publication&task=show&id=3732 [doi: 10.5593/sgemsocial2019v/1.1/s02.022]
28. Art. 35: GDPR Data protection impact assessment. General Data Protection Regulation (GDPR). URL: https://gdpr-info.eu/art-35-gdpr/ [accessed 2021-11-29]
29. Clarke R. Privacy impact assessment: its origins and development. Comput Law Sec Rev 2009 Jan;25(2):123-135. [doi: 10.1016/j.clsr.2009.02.002]
30. Oetzel MC, Spiekermann S. A systematic methodology for privacy impact assessments: a design science approach. Eur J Inform Syst 2019 Jan 28;23(2):126-150. [doi: 10.1057/ejis.2013.18]
31. Raab CD. Information privacy, impact assessment, and the place of ethics. Comput Law Sec Rev 2020 Jul;37:105404. [doi: 10.1016/j.clsr.2020.105404]

Abbreviations

CORE: Connected and Open Research Ethics
DHC-R: Digital Health Checklist for Researchers
HIPAA: Health Information Portability and Accountability Act
IRB: institutional review board
REC: research ethics committee
ReCODE: Research Center for Optimal Digital Ethics

Edited by G Eysenbach; submitted 31.10.20; peer-reviewed by I Schiering, C Calyx; comments to author 08.12.20; revised version received 21.12.20; accepted 05.11.21; published 22.12.21.

Please cite as:
Bartlett Ellis R, Wright J, Miller LS, Jake-Schoffman D, Hekler EB, Goldstein CM, Arigo D, Nebeker C
Lessons Learned: Beta-Testing the Digital Health Checklist for Researchers Prompts a Call to Action by Behavioral Scientists
J Med Internet Res 2021;23(12):e25414
URL: https://www.jmir.org/2021/12/e25414
doi:10.2196/25414
PMID:34941548

Viewpoint

Social Media and the Transformation of the Physician-Patient Relationship: Viewpoint

Ella M E Forgie[1], MA; Hollis Lai[2], PhD; Bo Cao[3], PhD; Eleni Stroulia[3], PhD; Andrew J Greenshaw[4], PhD; Helly Goez[5], MD

[1]Department of Anthropology, University of Alberta, Edmonton, AB, Canada
[2]Department of Dentistry, University of Alberta, Edmonton, AB, Canada
[3]Department of Computing Science, University of Alberta, Edmonton, AB, Canada
[4]Department of Psychiatry, University of Alberta, Edmonton, AB, Canada
[5]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Helly Goez, MD
Department of Pediatrics
University of Alberta
Rm 19 Glenrose Rehabilitation Hospital
10230 111 Ave NW
Edmonton, AB, T5G 0B7
Canada
Phone: 1 780 492 4119
Email: goez@ualberta.camd

Abstract

As many as 80% of internet users seek health information online. The social determinants of health (SDoH) are intimately related to who has access to the internet and health care as a whole. Those who face more barriers to care are more likely to benefit from accessing health information online, assuming the information they are retrieving is accurate. Virtual communities on social media platforms are beginning to serve as venues for seeking health information online because peers have been shown to influence health behavior more than almost anything else. As a positive mediator of health, social media can be used as a direct or indirect mode of communication between physicians and patients, a venue for health promotion and health information, and a community support network. However, false or misleading content, social contagion, confirmation bias, and security and privacy concerns must be mitigated to realize the full potential of social media as a positive mediator of health. This paper presents the shifting dynamics of how such communities are affecting physician-patient relationships. With the intersections between the SDoH, social media, and health evolving, physicians must take into consideration these factors when establishing their relationships with patients. We argue a paradigm shift in the physician-patient relationship is warranted, one where physicians acknowledge the impacts of the SDoH on information-seeking behavior, recognize the positive and negative roles of social media as a mediator of health through the lens of the SDoH, and use social media to catalyze positive changes in the physician-patient relationship. We discuss how the physician-patient relationship must evolve to accommodate for the ever-increasing role of social media in health and to best use social media as a tool to improve health outcomes. Finally, we present a fluid and multicomponent diagram that we believe will assist in framing future research in this area. We conclude that it is ineffective and even counterproductive for physicians to ignore the relationship between social media, the SDoH and health, their impact on one another, and the effect it has on designing the medical encounter and the delivery of care under the definition of precision medicine.

(J Med Internet Res 2021;23(12):e25230) doi:10.2196/25230

KEYWORDS

social media; social determinants of health; precision medicine; patient care

Introduction

Precision medicine, a comprehensive approach to patient care that takes into consideration genetics and the social determinants of health (SDoH) when diagnosing and treating disease [1,2], is prevailing among the priority areas for research in Canada and should be considered the standard for patient care worldwide [3]. The SDoH include socioeconomic factors that affect health

and well-being, such as income, education, and employment [4]. Such factors are affected by societal systems of oppression and are intimately linked to an individual's position in society relative to the societal status quo. As neither genetics nor the SDoH alone can indicate health status, the two must be viewed as intrinsically and inextricably linked to one another and thus to an individual's overall health and well-being [5]. By definition, a precision health care approach should be both personalized to the patient and humanistic. With a multitude of health factors that can influence and are influenced by the SDoH, they should be considered by physicians when managing their patients' health concerns. However, many SDoH and barriers to care issues continue to be overlooked by physicians [6-8].

Research has shown that as many as 80% of internet users seek health information online [9,10]. Social media platforms, internet-based user-driven community platforms for communication and sharing user-created content [11], have flourished as virtual communities where people exchange information and opinions, and seek support and advice from peers [12-14]. Such platforms include blogs (eg, WordPress), wikis (eg, Wikipedia), social bookmarking (eg, Reddit), social network sites (eg, Facebook), status update services (eg, Twitter), virtual world content (eg, Minecraft), and media-sharing sites (eg, YouTube) [11]. Specifically, health information seeking is among the most popular online activities with diet/nutrition, physical activity, signs and symptoms, treatment, and public health interventions as some common examples [8-10,12,13,15]. Social media has been studied in attempts to improve health outcomes with mixed results, but it is not yet clear what aspects of successful interventions precipitated their success [16]. Likewise, the reasons for failure of unsuccessful interventions, which led to worsened health outcomes or increased health inequity, are not clear. Thus, social media has the potential to positively moderate patient health, but how it can be used correctly to maximize benefit and minimize repercussions for patients remains to be seen.

In part related to the accepted importance of understanding patients' voices and perspectives, it has been implied that social media use should also be taken into consideration by physicians as a determinant of health [6]. Although health care practitioners remain the principal trusted authority for health information, the SDoH of the patient—including accessibility barriers such as geography, cost, and time—result in patient preference for online searching over in-person consultations [12]. Due to its social connectedness, social media is one of the preferred venues for obtaining health information and community support, and peers have been shown to influence health behaviors more than almost anything else [14]. There are questions as to the accuracy and unbiased nature of online health information, especially that which is disseminated on user-created content platforms such as social media [8,17,18]. Further, an individual's internet competency does not necessarily equate to their medical literacy. However, whether or not social media is a net positive or negative mediator of health, as we will argue, it undoubtedly affects individual health status in unique and substantial ways and thus cannot be ignored by physicians.

In this paper, we examine the complex relationship between social media and the SDoH. We propose that social media's effects on health should be considered as part of the standard of care. Moreover, we argue that a paradigm shift in the physician-patient relationship is warranted, one where physicians acknowledge the impacts of the SDoH on information-seeking behavior, recognize the positive and negative roles of social media as a mediator of health through the lens of the SDoH, and use social media to catalyze positive changes in the standard of care. We suggest that only by broadening our understanding of the intimate linkages between social media and the SDoH, and incorporating it as part of patient care, can the gap be bridged effectively to deliver a vision of precision medicine that is inherently and sufficiently personalized.

The Impacts of the Social Determinants of Health on Information-Seeking Behavior

The Government of Canada recognizes 12 determinants of health, including but not limited to biology and genetic endowment, childhood experiences, physical environment, and access to health services [4]. Not all determinants of health are SDoH. Rather, the SDoH are those that focus on social and economic factors such as race, income and social status, education and literacy, and employment and working conditions. The World Health Organization's Commission on the Social Determinants of Health has defined SDoH as "the conditions in which people are born, grow, live, work and age" [19].

It is imperative to consider who has access to the internet when discussing the demographics of social media users, as social media users are by definition a subset of internet users. Access to the internet in general, a requisite for accessing social media, is related to factors underlying the SDoH. This relationship between internet access and the SDoH influences what information individuals are accessing and how they are accessing it [20,21]. This, in turn, affects their use of digital tools such as social media, which encompasses a wide array of websites and applications. Nearly 60% of the global population have access to the internet [22]. Nevertheless, a digital divide exists, and although internet use increases every year, this digital divide highlights disparities in access for underserved populations, especially those in low- and middle-income countries (LMIC) [14,22,23].

Age, an SDoH, is also a factor contributing to the differentiated use of social media; older adults tend to search for health information online significantly less than younger age cohorts such as Generation Y (1977-1990) because they lack prior internet experience and thus possess lower internet competency [8,14,24,25]. In addition, evidence indicates that education, another recognized SDoH, and specifically higher education correlates with an increased likelihood of searching for health information online as does identifying as female [20,26].

Overall, people of low socioeconomic status or from LMIC, older adults, and less educated individuals face the most substantial barriers to accessing the internet in general and social media in particular. Unfortunately, this is the same population who stand to benefit the most, healthwise, from having access to health information online [14]. The apparent need for internet

access in specific SDoH segments reinforces clearly the statement delivered by the United Nations Human Rights Council that access to the internet is a basic human right [27].

Social Media's Potential as a Positive Mediator of Health

Accessibility barriers to in-person care are currently unacceptably high [14,22,23], and methods of communication between physicians and patients remain strictly in traditional formats [7]. Despite enormous advancements in internet technology and virtual communication, patients are still required to appear in-person at a physician's office to communicate with a physician and receive care. To do so requires substantial privilege on the part of the patient, such as having the time and resources to attend an appointment. Recently, due to the COVID-19 pandemic, there has been an acute increase in emphasis on providing more virtual care. However, this is still quite limited and "temporary" as "there isn't a lot of readiness" for using virtual health care [28]. Although methods of communication between physicians and patients are "at the heart of healthcare" [7], they are traditionally designed using practices that appear to lag substantially behind modern methods, in part due to regulatory bodies' regulations and practice guidelines [28]. Although patients tend to seek health information online for a number of reasons, the majority state that they would prefer to obtain this information from health care practitioners, but it is simply not within their means to do so due to an inability to access care as a result of economic, social, cultural, or physical barriers [23,29].

Social media is largely free, easy to access from multiple geographical locations, and considered by patients to be "more convenient, timely, cost-effective" [8]. In some cases, receiving care in person may be impeded by stigma (eg, for treatment of mental health or sexually transmitted infections). In that regard, social media is reported by users as "privacy protective, and [less embarrassing]" [8] as a venue for seeking health information. Such perceptions are not always accurate, as will be discussed in the following section, and they by no means represent all social media users [30]. Further, social media is inclusive; provides a sense of solidarity, hence enhancing the attribute of community support; and grants a greater perceived sense of control to patients over their own health [29].

Patients have reported that information received from physicians during in-person consultations was "not clear, satisfactory, or conductive for asking additional questions" [31]. This may be one of the reasons that patients turn to social media for health information. When this occurs, they report feeling more knowledgeable, confident, and empowered in their abilities to communicate with health care practitioners [6]. Patient satisfaction (the perceived standard of care) is thus improved when social media is used as a tool for obtaining health information.

Perhaps equally important to decreasing accessibility barriers and increasing patient access to health information, social media also acts as a support network [12,13]. One study found that support networks are a preferred venue for obtaining health information, second only to physicians [12]. Support networks, including those facilitated by social media, are linked to the SDoH [32-34] and substantially affect health behaviors [35]. This is useful for health promotion and health outcomes, as people may be persuaded to partake in positive health activities such as healthy diet, exercising [36-38], and receiving their annual flu shot if their peers have posted publicly about participating in these activities [39]. In sum, social media has the ability to mitigate the SDoH that result in limited physical accessibility, enhance personal confidence, and empower patients' communication with their physicians.

Social Media's Potential as a Negative Mediator of Health

For all the aforementioned benefits, social media is not exclusively a positive mediator of health [29,40]. There are notable problems with obtaining health information on social media; direct implications of social media use on health; and a problematic, homogeneous, "single story" narrative that is presented. It is important to acknowledge the negative and potentially dangerous effects of social media on health to reconcile them.

Social media tends to present information of questionable credibility, and it is oftentimes sponsored by a potentially biased entity [8]. Occasionally, online health information is entirely false [17]. In addition, although it may provide patients with a sense of privacy in comparison to discussing stigmatized topics in person with physicians, there are notable concerns about anonymity and privacy when obtaining health information online [29]. Confirmation bias (ie, the tendency to seek and trust information that is belief consistent and to discredit information that is belief inconsistent) is another well-documented danger associated with seeking health information online [18]. Personalization algorithms on social media platforms can further polarize the information available to patients though "recommended" or "suggested" content [13,41]. This content is automatically sourced based on previous social media activity, and it is presented to the user whether they are seeking it or not. Antivaccination content is a particularly good example, as parents who search for vaccine information online are more likely to hold antivaccination beliefs and possibly be active on similar communities of interest [10,18].

Emerging literature is increasingly documenting the direct impacts of social media on health [13,35]. Social media has been found to promote a sedentary lifestyle, increase self-isolation, decrease quantity and quality of sleep, and negatively impact mental health [13]. Social contagion, a recently documented phenomenon, depicts the contagious nature of certain noncommunicable health conditions over social media, including obesity and emotions such as happiness, anxiety, and depression [35]. Social influence and peer recommendations may substantially alter a person's health behaviors as well, likely due to susceptibility to peer pressure, desire to belong to a group or feel supported, and perception of credibility of the recommendations [14,35,42]. A vicious cycle between anxiety and online health information seeking is another documented phenomenon, where high levels of anxiety are associated with

online health information seeking, the findings of which further increases anxiety [43].

In 2009, Chimamanda Ngozi Adichie delivered a now renowned TEDtalk about the dangers of a "single story" narrative. Single stories emerge when only one narrative about a group of people is widely shared and accepted. It then becomes assumed that all members of said group shared the same experience, one that aligns with the single story. Recently, this concept was extended to stories of health experiences shared on social media [40]. An anthropological study found that single stories of health experiences (surrogacy, in this particular example) are disseminated widely across social media platforms. Not only do these single stories ignore variation in individual experiences and personal SDoH, but they also tend to hold little truth at all. Any deviation from the single story on social media, including any disclosure of personal experiences, is discredited and rejected, dehumanizing those who do so. Physicians must be made aware of the homogeneity of information on social media and how this may affect their patients' perceptions of health and health experiences. Further, patients using social media as an online community wherein they may communicate with others with similar health conditions must be empowered to tell their own story while acknowledging that it may not align with the stories of others. Subjective recounts of disease are not necessarily misinforming so long as patients understand the difference between individual perception and experience of disease versus etiology and treatment of disease, the latter of which necessitates a degree of prerequisite medical knowledge that goes above and beyond digital literacy [44,45].

Using Social Media to Catalyze Positive Changes in the Physician-Patient Relationship

Although the positive role of social media as a mediator of health is ample (ability to improve methods of direct and indirect communication between physicians and patients, increase access to health information, and foster a community support network), it is also imperative to address and mitigate the negative aspects of social media, such as security and privacy concerns. Work must be done on the part of the physician and the patient to shift the physician-patient relationship toward one that is inclusive of the role that social media plays in health and that uses social media as a tool to promote health and well-being.

During in-person consultations, physicians must be open to discussing the roles of social media as a mediator of health. In a recent interview, patients perceived health care workers to be in "overt or tacit opposition" [6] to any mention by patients of health information retrieved online. Such mentality reflects the traditional physician-patient relationship where patients hold little autonomy over their own medical journey [23]. Take, for example, a scenario wherein a patient, Cindy, a 24-year-old veterinary technician, discloses to her physician concerns around the COVID-19 vaccines. Cindy is a first-generation immigrant living in a multigenerational home and is worried that she may become contagious with the virus immediately after immunization, posing a health risk to her older adult immunocompromised grandmother. If her grandmother were to fall ill, Cindy fears that her family would not be able to afford the necessary care. Therefore, Cindy feels it is best for her family not to get the vaccine. For the physician to mitigate these concerns, it is imperative that they attempt to discern from where these anxieties originate. To write off Cindy's beliefs simply because they are biomedically unfounded and derived from information sought on social media could be counterproductive in protecting or improving her health.

Further, social media may be used as an indirect line of communication between physicians and patients, if used as a venue for health knowledge dissemination and translation for the purpose of health promotion [46,47]. To elaborate, physicians or health care experts may communicate with their patients on social media via public posts about health-promoting behaviors they recommend and providing them with trusted links for further information. In addition, physicians may refer their patients to patient-driven online advocacy groups such as the Light Collective, a nonprofit organization that, among other objectives, works to shine a light on privacy breaches and so-called "bad data-sharing" [30]. This would aid in decreasing misinformation and diversifying the pervasive single-story narratives that encapsulate many health conditions. We recognize that it is not only up to the physician to identify and quell misinformation on social media; it is largely up to the platform itself to create regulatory policies and practices—or better yet, algorithms to identify problematic posts [48]—to minimize false or misleading content and to ensure that it is not being amplified in "recommended" or "suggested" content. In the previously discussed scenario, it is possible that Cindy would not have developed such concerns over the COVID-19 vaccines had she viewed credible biomedical data, or conversely, had she not viewed biomedically unfounded data, on social media. This is especially likely if Cindy is passively consuming data (ie, not actively seeking data).

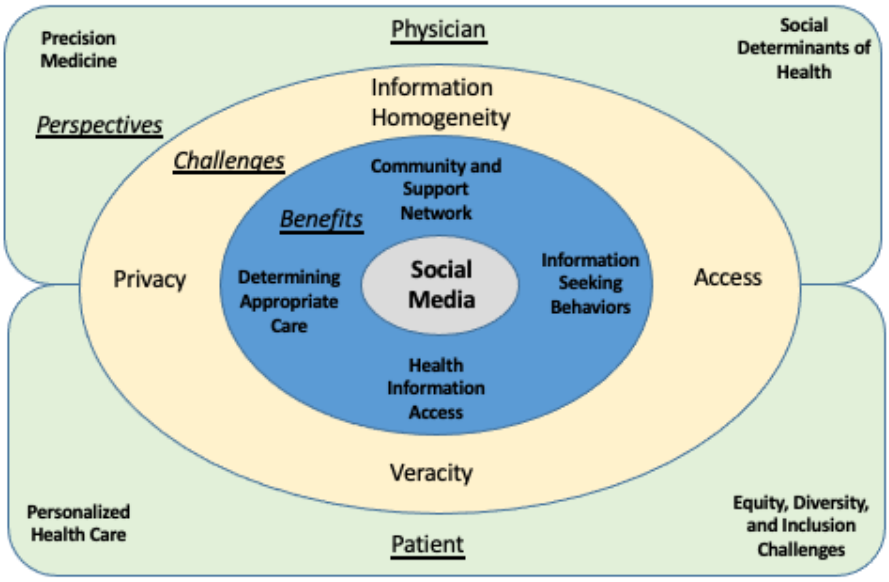
A patient's level of education and literacy as an SDoH impacts their level of medical literacy [49], which is only as good as the information they have acquired; it may be rendered obsolete if the knowledge they have gained is inaccurate. Therefore, patients should feel empowered to ask questions during their in-person consult or through anonymous or confidential online forums wherein knowledgeable health care practitioners provide answers. Although it is important for physicians to provide health care information to patients in terms that they are able to understand, we posit that it may also be beneficial to provide patients with a list of the formal medical terms associated with their condition. This may improve the caliber of search results they find online and reduce the likelihood of conflating different conditions or symptoms based on the colloquial descriptions given to them by their health care provider. Finally, patients should be encouraged to approach social media platforms in the context of health with curiosity and skepticism, embracing the community and solidarity aspects they may provide to those facing similar health problems while ensuring that objective medical data are not trusted until substantiated by scholarly sources or health care practitioners. For Cindy, a list of search terms will allow her to return home and actively seek reputable information on the COVID-19 vaccines. On the contrary, a lack

of take-home information may result in Cindy searching what she knows—"contagious after COVID-19 vaccine"—which may elicit misguiding results that confirm her bias. A brief but thorough discussion with her physician about the dangers of passively consuming bad data on social media may also prevent such encounters in the future.

Summary

In this paper, we reviewed the complex, multifaceted, and dynamic intersections of social media and SDoH in the era of precision medicine. We argued for its inclusion as part of routine patient care and demonstrated the potential of social media to be used as a positive mediator of health so long as its negative mediation effects are minimized. Further, we discussed how the physician-patient relationship must shift to accommodate for the ever-increasing role of social media in health and best use social media as a tool to improve health outcomes. These intricacies warrant further research. However, to conceptualize our study into a framework of understanding and development in this area, we present Figure 1 that we believe captures the potential challenges and areas of contention in bringing social media into precision medicine.

Figure 1. A fluid and multicomponent diagram to assist in framing future research around social media and the transformation of the physician-patient relationship.



1. The benefits of social media use in health care are listed in the first ring surrounding social media. Social media can act as a support network, promoting information-seeking behaviors, allowing patients to determine their own appropriate care, and enabling health information access, all affected by and affecting the SDoH.
2. However, social media also brings challenges to promoting health, as is illustrated in the second ring. Although privacy and access to social media are common issues, the public also faces the issue of veracity of information. Further, patients face an information homogeneity problem in the form of social media's single story, which does not account for diversity of experiences embodied by the SDoH.
3. Physicians and patients have different perceptions of how social media may be used in relation to health care, largely due to miscommunication. Although the public expects to receive personalized health care, physicians deliver their version of this as precision medicine. In addition, although the public perceives societal issues as equity, diversity, and inclusion challenges, physicians understand these issues as they relate to health and categorize them as the SDoH. This is not an issue of incongruity but one of language. Communication barriers must be acknowledged to be overcome.
4. Future research in this domain needs to recognize the complex dynamics of how social media interacts with the SDoH to develop solutions that can comprehensively improve the delivery of health care in the future.

Conclusion

As the role of social media in health evolves, new directions of research are needed to better understand the impacts of social media on health and inform physicians on how it can be integrated as part of patient care. Our discussion posits that health and social media are intimately linked through the prism of the SDoH and that this linkage is only amplifying over time. We argue that it is thus ineffective and even counterproductive for physicians to ignore this relationship and the impact it has on designing the medical encounter and the delivery of care. For physicians to deliver the highest standard of care under the definition of precision medicine, the complex interaction between social media and the SDoH and their impact on one another must be taken into consideration.

Acknowledgments

This study was partially supported by the AI4Society signature area of the University of Alberta.

Authors' Contributions

EMEF drafted and critically reviewed the manuscript. HL, BC, ES, AJG, and HG conceived the project and critically reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. What is precision medicine? MedlinePlus. URL: https://ghr.nlm.nih.gov/primer/precisionmedicine/definition [accessed 2020-09-01]
2. Vogenberg F, Isaacson Barash C, Pursel M. Personalized medicine: part 1: evolution and development into theranostics. P & T 2010 Oct;35(10):560-576 [FREE Full text] [Medline: 21037908]
3. Research in priority areas. Canadian Institutions of Health Research. 2020. URL: https://cihr-irsc.gc.ca/e/50077.html [accessed 2020-09-01]
4. Social determinants of health and health inequalities. Government of Canada. URL: https://www.canada.ca/en/public-health/services/health-promotion/population-health/what-determines-health.html [accessed 2020-09-01]
5. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129 Suppl 2:19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
6. Benetoli A, Chen T, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: 10.1016/j.pec.2017.08.015] [Medline: 28882545]
7. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
8. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
9. Fox S, Fallows D. Internet health resources. Pew Research Center. 2003. URL: https://www.pewresearch.org/internet/2003/07/16/internet-health-resources/ [accessed 2020-09-09]
10. Meleo-Erwin Z, Basch C, MacLean SA, Scheibner C, Cadorett V. "To each his own": Discussions of vaccine decision-making in top parenting blogs. Hum Vaccin Immunother 2017 Aug 03;13(8):1895-1901 [FREE Full text] [doi: 10.1080/21645515.2017.1321182] [Medline: 28481675]
11. Dewing M. Social Media: An Introduction. Ottawa, ON: Library of Parliament; 2010:1-5.
12. Criss S, Woo Baidal JA, Goldman R, Perkins M, Cunningham C, Taveras E. The role of health information sources in decision-making among hispanic mothers during their children's first 1000 days of life. Matern Child Health J 2015 Nov;19(11):2536-2543 [FREE Full text] [doi: 10.1007/s10995-015-1774-2] [Medline: 26122256]
13. Goodyear VA, Armour KM, Wood H. Young people and their engagement with health-related social media: new perspectives. Sport Educ Soc 2019;24(7):673-688 [FREE Full text] [doi: 10.1080/13573322.2017.1423464] [Medline: 31814804]
14. Lau AYS, Siek KA, Fernandez-Luque L, Tange H, Chhanabhai P, Li SYW, et al. The role of social media for patients and consumer health. Yearb Med Inform 2018 Mar 06;20(01):131-138. [doi: 10.1055/s-0038-1638751]
15. Wong DK, Cheung M. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
16. 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. [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
17. Moukarzel S, Del Fresno M, Bode L, Daly A. Distance, diffusion and the role of social media in a time of COVID contagion. Matern Child Nutr 2020 Oct;16(4):e13025 [FREE Full text] [doi: 10.1111/mcn.13025] [Medline: 32343487]
18. Meppelink C, Smit E, Fransen M, Diviani N. "I was Right about Vaccination": confirmation bias and health literacy in online health information seeking. J Health Commun 2019;24(2):129-140. [doi: 10.1080/10810730.2019.1583701] [Medline: 30895889]
19. Sheiham A. Closing the gap in a generation: health equity through action on the social determinants of health. A report of the WHO Commission on Social Determinants of Health (CSDH) 2008. Community Dent Health 2009 Mar;26(1):2-3. [Medline: 19385432]
20. Myrick JG, Willoughby JF. Educated but anxious: How emotional states and education levels combine to influence online health information seeking. Health Informatics J 2019 Sep;25(3):649-660 [FREE Full text] [doi: 10.1177/1460458217719561] [Medline: 28728457]

21. Khoong EC, Le GM, Hoskote M, Rivadeneira NA, Hiatt RA, Sarkar U. Health information-seeking behaviors and preferences of a diverse, multilingual urban cohort. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S176-S183 [FREE Full text] [doi: 10.1097/MLR.0000000000001050] [Medline: 31095058]
22. Clement J. Global digital population as of July 2020. Statista. 2020. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2020-09-01]
23. De Jesus M. The impact of mass media health communication on health decision-making and medical advice-seeking behavior of u.s. Hispanic population. Health Commun 2013;28(5):525-529. [doi: 10.1080/10410236.2012.701584] [Medline: 22888787]
24. Cisu T, Mingin G, Baskin L. An evaluation of the readability, quality, and accuracy of online health information regarding the treatment of hypospadias. J Pediatr Urol 2019 Feb;15(1):40.e1-40.e6. [doi: 10.1016/j.jpurol.2018.08.020] [Medline: 30449679]
25. Kyriacou A, Sherratt C. Online health information-seeking behavior by endocrinology patients. Hormones (Athens) 2019 Dec;18(4):495-505 [FREE Full text] [doi: 10.1007/s42000-019-00159-9] [Medline: 31749117]
26. Ashkanani H, Asery R, Bokubar F, AlAli N, Mubarak S, Buabbas A, et al. Web-based health information seeking among students at Kuwait University: cross-sectional survey study. JMIR Form Res 2019 Oct 31;3(4):e14327 [FREE Full text] [doi: 10.2196/14327] [Medline: 31473592]
27. The promotion, protection and enjoyment of human rights on the Internet: draft resolution. United Nations Digital Library. 2016. URL: https://digitallibrary.un.org/record/845728?ln=en [accessed 2020-09-01]
28. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
29. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
30. Our mission. The Light Collective. URL: https://lightcollective.org/mission/ [accessed 2020-09-01]
31. Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of Health Information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
32. Stewardt M. Social network, social support, and health. CJNR 2009;41(3):7-9.
33. Abnousi F, Rumsfeld JS, Krumholz HM. Social determinants of health in the digital age: determining the source code for nurture. JAMA 2019 Jan 22;321(3):247-248. [doi: 10.1001/jama.2018.19763] [Medline: 30570663]
34. Wright K. Social networks, interpersonal social support, and health outcomes: a health communication perspective. Front Commun 2016 Oct 14;1(10):10. [doi: 10.3389/fcomm.2016.00010]
35. Berkman L, Glass T. Social integration, social networks, social supporthealth. In: Berkman LF, Kawachi I, editors. Social Epidemiology. Oxford, England: Oxford University Press; 2000.
36. So J, Kuang K, Cho H. Information seeking upon exposure to risk messages: predictors, outcomes, and mediating roles of health information seeking. Commun Res 2016 Nov 24;46(5):663-687. [doi: 10.1177/0093650216679536]
37. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
38. McGloin AF, Eslami S. Digital and social media opportunities for dietary behaviour change. Proc Nutr Soc 2015 May;74(2):139-148. [doi: 10.1017/S0029665114001505] [Medline: 25319345]
39. Liao Q, Fielding R, Cheung Y, Lian J, Yuan J, Lam W. Effectiveness and parental acceptability of social networking interventions for promoting seasonal influenza vaccination among young children: randomized controlled trial. J Med Internet Res 2020 Feb 28;22(2):e16427 [FREE Full text] [doi: 10.2196/16427] [Medline: 32130136]
40. Teman E. The power of the single story: surrogacy and social media in Israel. Med Anthropol 2019 Apr;38(3):282-294. [doi: 10.1080/01459740.2018.1532423] [Medline: 30457365]
41. Guo W, Tao Y, Li X, Lin X, Meng Y, Yang X, et al. Associations of internet addiction severity with psychopathology, serious mental illness, and suicidality: large-sample cross-sectional study. J Med Internet Res 2020 Aug 11;22(8):e17560 [FREE Full text] [doi: 10.2196/17560] [Medline: 32780029]
42. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
43. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
44. Holt K, Overgaard D, Engel LV, Kayser L. Health literacy, digital literacy and eHealth literacy in Danish nursing students at entry and graduate level: a cross sectional study. BMC Nurs 2020;19:22 [FREE Full text] [doi: 10.1186/s12912-020-00418-w] [Medline: 32308559]
45. Klecun E. Digital literacy for health: the promise of health 2.0. Int J Digital Literacy Digital Competence 2010;1(3):10. [doi: 10.4018/jdldc.2010070105]

46. Katz MS, Anderson PF, Thompson MA, Salmi L, Freeman-Daily J, Utengen A, et al. Organizing online health content: developing hashtag collections for healthier internet-based people and communities. JCO Clin Cancer Inform 2019 Jun;3:1-10 [FREE Full text] [doi: 10.1200/CCI.18.00124] [Medline: 31251658]
47. McClain C. Practices and promises of Facebook for science outreach: becoming a "Nerd of Trust". PLoS Biol 2017 Jun;15(6):e2002020 [FREE Full text] [doi: 10.1371/journal.pbio.2002020] [Medline: 28654674]
48. 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]
49. Larsen K. Health literacy. In: Walker PF, Barnett ED, editors. Immigrant Medicine. Amsterdam: Elsevier Inc; 2007:E-24.

Abbreviations

LMIC: low- and middle-income countries
SDoH: social determinants of health

Edited by G Eysenbach; submitted 23.10.20; peer-reviewed by P Aslani, M Afzal, L Salmi; comments to author 23.12.20; revised version received 25.01.21; accepted 26.10.21; published 24.12.21.

Please cite as:
Forgie EME, Lai H, Cao B, Stroulia E, Greenshaw AJ, Goez H
Social Media and the Transformation of the Physician-Patient Relationship: Viewpoint
J Med Internet Res 2021;23(12):e25230
URL: https://www.jmir.org/2021/12/e25230
doi:10.2196/25230
PMID:34951596

Original Paper

Clinical Efficacy and Psychological Mechanisms of an App-Based Digital Therapeutic for Generalized Anxiety Disorder: Randomized Controlled Trial

Alexandra Roy[1], BA; Elizabeth A Hoge[2], MD; Pablo Abrante[1], BSc; Susan Druker[3], MA; Tao Liu[4], PhD; Judson A Brewer[1], MD, PhD

[1]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States

[2]Department of Psychiatry, Georgetown University Medical Center, Washington, DC, United States

[3]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Department of Biostatistics, Brown University School of Public Health, Providence, RI, United States

Corresponding Author:
Judson A Brewer, MD, PhD
Department of Behavioral and Social Sciences
Brown University School of Public Health
121 South Main Street
Providence, RI, 02903
United States
Phone: 1 401 863 2826
Email: judson_brewer@brown.edu

Abstract

Background: Current treatments for generalized anxiety disorder (GAD) often yield suboptimal outcomes, partly because of insufficient targeting of underlying psychological mechanisms (eg, avoidance reinforcement learning). Mindfulness training (MT) has shown efficacy for anxiety; yet, widespread adoption has been limited, partly because of the difficulty in scaling in-person–based delivery. Digital therapeutics are emerging as potentially viable treatments; however, very few have been empirically validated.

Objective: The aim of this study is to test the efficacy and mechanism of an app-delivered MT that was designed to target a potential mechanism of anxiety (reinforcement learning), based on which previous studies have shown concern regarding feedback and the perpetuation of anxiety through negative reinforcement.

Methods: Individuals with GAD were recruited using social media advertisements and randomized during an in-person visit to receive treatment as usual (n=33) or treatment as usual+app−delivered MT (Unwinding Anxiety; n=32). The latter was composed of 30 modules to be completed over a 2-month period. Associated changes in outcomes were assessed using self-report questionnaires 1 and 2 months after treatment initiation.

Results: We randomized 65 participants in this study, and a modified intent-to-treat approach was used for analysis. The median number of modules completed by the MT group was 25.5 (IQR 17) out of 30; 46% (13/28) of the participants completed the program. In addition, the MT group demonstrated a significant reduction in anxiety (GAD-7) compared with the control group at 2 months (67% vs 14%; median change in GAD-7: –8.5 [IQR 6.5] vs –1.0 [IQR 5.0]; P<.001; 95% CI 6-10). Increases in mindfulness at 1 month (nonreactivity subscale) mediated decreases in worry at 2 months (Penn State Worry Questionnaire; P=.02) and decreases in worry at 1 month mediated reductions in anxiety at 2 months (P=.03).

Conclusions: To our knowledge, this is the first report on the efficacy and mechanism of an app-delivered MT for GAD. These findings demonstrate the clinical efficacy of MT as a digital therapeutic for individuals with anxiety (number needed to treat=1.6). These results also link recent advances in our mechanistic understanding of anxiety with treatment development, showing that app-delivered MT targets key reinforcement learning pathways, resulting in tangible, clinically meaningful reductions in worry and anxiety. Evidence-based, mechanistically targeted digital therapeutics have the potential to improve health at a population level at a low cost.

Trial Registration: ClinicalTrials.gov NCT03683472; https://clinicaltrials.gov/ct2/show/NCT03683472

(J Med Internet Res 2021;23(12):e26987) doi:10.2196/26987

KEYWORDS

anxiety; generalized anxiety disorder; worry; mindfulness; mHealth; digital therapeutics; mobile phone

Introduction

Background

Anxiety disorders are the most common class of mental illnesses, with a 31% lifetime prevalence [1]. This has already increased during the COVID-19 pandemic [2-4]. For example, in the United States, the Census Bureau reported that adults were more than 3 times more likely to screen positive for an anxiety disorder in 2020 than they were in 2019 (31% vs 8%), and a cross-sectional survey of people in China in 2020 reported the prevalence of generalized anxiety disorder (GAD) to be 35.1% [2]. A recent meta-analysis of 17 studies (N=62,000) found an average prevalence of anxiety of 32% during COVID-19 [5].

Treatment for Anxiety

However, anxiety, particularly GAD, is difficult to treat. Current practice guidelines recommend pharmacological and psychological interventions [6], but most patients favor psychotherapy over medications [7]. Benzodiazepines have the risk of tolerance and addiction; the United Kingdom National Institute for Health and Care Excellence guidelines state that benzodiazepines "should not be used routinely to treat anxiety disorders" [6]. First-line treatment for anxiety, such as selective serotonin reuptake inhibitors, can have limitations, including delayed patient responses and adverse effects (eg, gastrointestinal and sexual) [8,9]. The number needed to treat (NNT) for selective serotonin reuptake inhibitors is 5.2; one needs to treat >5 individuals to see a significant response in 1 individual [10].

Cognitive behavioral therapy (CBT), the most commonly used and researched psychological intervention for anxiety, has been associated with reduction in symptoms using measures such as the Beck Anxiety Inventory (small to medium effect sizes), and yet is typically delivered in-person [11]. Given the growing need and shortage of trained therapists [12,13], new mechanistically based treatments that can be delivered at scale and a distance are needed for those unable to attend in person because of barriers such as living in a resource-limited area or unwillingness to go to a mental health clinic because of stigma. In addition, treatments that target more recent mechanistic underpinnings of anxiety are needed.

Psychological Mechanisms of Anxiety

From a theoretical standpoint, reinforcement learning mechanisms have been suggested to drive anxiety disorders [14,15]. Worry is widely regarded as the central defining feature of GAD and has been shown to be triggered as an avoidance reaction to emotional experiences [16], thus learned as "a negatively reinforced avoidant behavior" [17]. Recent research has linked reinforcement learning with biological mechanisms and clinical symptoms [18]. Worry represents an attempt to engage in mental problem solving on an issue with an uncertain outcome [19]. Although worry is unpleasant, the immediate emotions that are avoided by focusing on worry, such as fear, are often perceived as *more* unpleasant [20,21]. Therefore, aversive stimuli can trigger worry as an avoidant behavior, which then becomes habitual [22]. Theoretically and mechanistically, worry is learned and reinforced in a manner similar to other types of operantly conditioned behaviors [14]. With cycles of reinforcement learning, when triggered by feelings of anxiety, individuals with GAD learn a maladaptive thinking style that uses worry as a focus on the future rather than the present. As described by Borkovec et al [14,15], when worry becomes habitual, its negative reinforcement pathway can spiral out of control; when the negative emotional experience of worry rises to the same level as unpleasant emotional states that trigger it, it may become its own trigger for more anxiety, which leads to more worry directly driving anxiety *habit loops* (Gao et al, unpublished data, 2021).

In conventional frameworks, anxiety is conceptualized as an overestimation of danger and an underestimation of one's ability to cope with it [23]. Cognitive therapies aim to interrupt the cycle of worry by replacing maladaptive cognitions with more constructive ones. For individuals who cycle into anxious worry to a degree that impairs prefrontal cortical function and the ability to use cognitive therapies [24], other strategies are needed. New treatments such as mindfulness training (MT) have shown promise in efficacy and cost, with effect sizes rivaling current treatments [25]. Furthermore, cross-sectional studies of nonclinical populations have suggested a mediating role of worry in the effects of MT on anxiety [26].

Mindfulness

Mindfulness can be defined as the awareness that arises when paying attention to the present moment on purpose and nonjudgmentally [27]. The attitudinal quality of not judging and allowing experience to unfold with curiosity targets maladaptive reinforcement learning by helping individuals to simply observe repetitive cycles of perseverative worry rather than to habitually react and reinforce them [28]. MT has been found to mechanistically break key links in the reinforcement pathway for other habitual behaviors such as smoking and emotional eating [29,30], with concomitant changes in related brain regions predicting clinical outcomes [31]. Worry has been shown to activate brain networks associated with self-referential processing, such as the default-mode network, suggesting that the more one is *caught up* in perseveration about uncertain events, the more this network is activated [32,33]; meditation has been shown to directly deactivate these brain regions [34], and neurophenomenological studies suggest that this may be because of the ability to observe thoughts and emotions rather than being caught up in them [35-37]. Specific to GAD, MT has been associated with changes in the fronto-limbic brain regions involved in emotion regulation with simultaneous improvements in reported symptoms [38].

App-Delivered MT

Regarding treatment delivery, mindfulness-based interventions, such as mindfulness-based stress reduction, are generally delivered in a group format. However, concerns remain

regarding the scalability of in-person–delivered treatment [39]. Digital therapeutics (ie, app-delivered interventions) have garnered much attention as a new modality that can deliver high-fidelity treatment at scale and low cost. However, empirically tested apps are not widely used, and widely used apps do not have an evidence base (<0.05%)—a clear "digital research practice gap" [40]. To date, only 1 study has reported the clinical efficacy of a digital therapeutic specifically for people diagnosed with GAD (showing that CBT may be effective when delivered in this format) [41].

To address the digital research practice gap, we designed an app-delivered digital therapeutic program for anxiety that mechanistically targets reinforcement learning using MT to help individuals identify habitual worry thinking patterns and learn not to habitually react to unpleasant emotions (ie, break the worry cycle). In a single-arm study of anxious physicians, we found preliminary evidence for its utility in reducing anxiety (57% reduction in GAD-7 scores after 3 months) [42]. However, randomized controlled studies are required to determine the efficacy and mechanisms of action. We tested the following hypotheses in a randomized controlled trial of individuals with GAD: (1) app-delivered MT would show superior efficacy in reducing anxiety and worry than standard treatment; (2) increases in nonreactivity would mediate reductions in worry; and (3) reductions in worry would mediate reductions in anxiety.

Methods

Study Overview and Participants

We used a parallel-group randomized controlled trial design with analyses and outcome measures preregistered at ClinicalTrials.gov. Individuals were recruited using Facebook advertisements and screened for eligibility via a phone call by the project director (AR). Textbox 1 details the inclusion and exclusion criteria. These criteria were chosen to mimic, as closely as possible, real-world clinical situations while minimizing potential confounders (eg, a recent change in medication dose may mask treatment effects). Eligible participants attended an in-person research visit at Brown University, where they underwent informed consent procedures with the project director before enrolling in the study. Participants were provided Amazon gift cards worth up to US $80 to complete the questionnaires. This study was approved by the Institutional Review Board of Brown University.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Score ≥10 on the Generalized Anxiety Disorder (GAD) 7-item scale, which is suggestive of a diagnosis of GAD (sensitivity and specificity of 89% and 82%, respectively) [43]
- Owns a smartphone
- Willingness to receive check-in calls
- 18 years or older

Exclusion criteria

- Dose changes for any psychoactive medication in the last 2 months
- As-needed use of benzodiazepines or hypnotic sleep aids
- History of bipolar, schizophrenia, schizoaffective, or another psychotic disorder
- Significant medical condition that would impact the ability to complete study tasks
- Cohabitation with someone already enrolled in the study
- Previous use of other related apps

After enrolling participants in the study, the project director, who had previously undergone training and was supervised by a psychiatrist, conducted an in-person, abbreviated version of the Mini-International Neuropsychiatric Interview to confirm a diagnosis of GAD, along with the assessment of other potential comorbid disorders, such as major depressive episode, panic disorder, agoraphobia, social anxiety disorder, obsessive compulsive disorder, and posttraumatic stress disorder [44]. Participants were then asked to complete a web-based questionnaire via Qualtrics [45], which included demographic and self-report items. Participants in both groups received questionnaires via email 1 and 2 months after treatment initiation. Upon completion of the final questionnaire, the treatment as usual (TAU) group received instructions on how to download and install the app.

Randomization and Blinding

After completing the baseline questionnaire, participants were given a sealed, opaque envelope (prepared, reviewed, and sealed by individuals independent of the study team) by the project director that contained their group assignment: TAU+app-delivered MT or TAU. The 1:1 randomization scheme was generated by an independent statistician with variable block sizes of 4 and 6. Team members who randomized the participants and carried out the study procedures did not perform the study analyses. The principal investigator and the statistician who conducted the statistical analysis were blinded to the group allocation until all analyses were complete.

Intervention

The app-delivered MT program, Unwinding Anxiety (version 2), is a Health Insurance Portability and Accountability Act

compliant digital therapeutic that teaches individuals to understand how anxious worry develops and perpetuates through reinforcement learning and how to bring mindful awareness to moments of stress and worry so that they can observe feelings of anxiety rather than perpetuate reactive worry thinking. This process helps individuals *unlearn* or extinguish worry at the core mechanistic level. This experiential education is delivered via a smartphone-based platform, which includes a progression through >30 daily modules of brief didactic and experience-based MT (videos and animations approximately 10 minutes per day; Multimedia Appendix 1), app-triggered check-ins, user-initiated guided meditations (5-15 minutes), and brief (30 seconds) on-demand mindfulness exercises to help disrupt anxiety cycles in vivo (Textbox 2; Multimedia Appendix 2). The content for this intervention was developed based on a combination of clinical experiences and previously developed, in-person and app-delivered MT protocols for habit change that have yielded clinically meaningful outcomes (eg, smoking and overeating) [28-30,46-49]; an open-label pilot study of the app demonstrated a 57% reduction in GAD-7 scores in anxious physicians [42].

Textbox 2. Overview of Unwinding Anxiety themes and content.

Modules 1-7 (goals, curiosity, reinforcement learning, body scan, and self-monitoring)

- Set goals and introduce how habits are formed around worry (eg, reinforcement learning and distraction); introduce curiosity to foster the nonjudgmental aspects of mindfulness and basic mindfulness practices, including the body scan; and unpack worry and fear both from a brain and behavior perspective.

Modules 8-14 (noting practice; RAIN [recognize, accept, investigate, and note]; barriers to change; and reinforcement of concepts)

- Introduce how to mindfully work with worry cues and affective states using RAIN (recognize, accept, investigate, and note what emotions feel like as they arise and pass away); build on basic mindfulness using noting practice (the N of RAIN) during everyday life; and introduce additional animations to reinforce mindfulness concepts that show how we feed our anxiety by worry thinking and distraction.

Modules 15-21 (noting practice [continued from previous modules]; RAIN [continued from previous modules]; thinking versus knowing; and *un*resistance)

- Reinforce noting practice and continue to train and support self-kindness; specifically address the difference between trying to think our way out of uncertainty (or anxiety) and resting in a kind, curious awareness of it; and focus on not resisting experience and not getting tripped up by worry thinking.

Modules 22-30 (noting practice [continued from previous modules]; RAIN [continued from previous modules]; and working with uncertainty and change)

- Help individuals reflect on their own evidence base for working with worry to solidify their shift from reactivity to mindfully being with emotions as a new habit.

Modules 30 and onward (reinforcing concepts via *theme weeks* and individual customization via *personal week*)

- >8 themed weeks and unlimited personalization of content by picking modules to develop a custom week for review.

Intervention Orientation and Engagement

Individuals randomized to TAU+app-delivered MT were assisted with the installation of the app on their smartphone and the reviewing of the features. They were instructed to complete 1 module per day over the subsequent 30 days at a time of their choice. In addition, the intervention would check in with them 3 times throughout the day (this could be modified by the user) and offer brief mindfulness exercises. Participants were encouraged to use other app features but were informed that this was not a requirement for the study. The project director sent check-in messages on days 3, 7, 14, and 21 from treatment initiation to help mitigate technical difficulties and encourage engagement. Specifically, participants were asked "how things were going with the app since the last time they received a check-in." If the participant expressed difficulties, efforts were made to resolve the problem.

TAU Condition

As part of TAU, participants were asked to continue the standard care set forward by their clinician or clinicians. This could include pharmacological treatment or psychotherapy. Participants were also provided with a list of local resources.

Outcome Measures

The primary outcomes were changes in anxiety, as measured by the GAD-7, and emotional reactivity at 2 months after treatment initiation. Secondary outcomes included changes in worry, as measured by the Penn State Worry Questionnaire (PSWQ) and interoceptive awareness.

GAD 7-Item

GAD-7 is a validated 7-item questionnaire used clinically to screen for probable diagnosis of GAD (sensitivity of 89% and specificity of 82%; high internal consistency, with Cronbach α=.92) and track symptom severity [43]. Individuals are asked, "in the last week, how often have you been bothered by the following problems" with prompts such as *feeling nervous, anxious, or on edge* and *trouble relaxing* [50]. The scale ranges from 0 (*not at all sure*) to 3 (*nearly every day*), with scores ranging from 0 to 21 [50]. Total scores of 5, 10, and 15 serve as cut-off points for mild, moderate, and severe anxiety, respectively; therefore, remission is a score of ≤4 [50]. The

minimal clinically important difference for GAD-7 was 3.8; the clinically relevant change was ≥4 points [51]. GAD-7, which is highly correlated with the Hamilton Anxiety Scale (r=0.852), was used based on its *real-world* utility, as it is the most commonly employed tool in primary care and other outpatient settings and has been incorporated into most large-scale electronic medical record systems [52].

Five Facet Mindfulness Questionnaire Nonreactivity Subscale

The nonreactivity subscale is composed of 7 questions from the 39-item Five Facet Mindfulness Questionnaire (FFMQ) with acceptable internal consistency (Cronbach α=.75) [53]. It is validated for use independently and assesses nonreactivity to inner experience [54]. Individuals are asked questions about what is generally true for them on a scale from 1 (*never or very rarely true*) to 5 (*very often or always true*) [53]. Examples include "I perceive my feelings and emotions without having to react to them" and "When I have distressing thoughts, I feel calm soon after" [53]. Scores range from 7 to 35, with higher scores indicating an increase in nonreactivity.

Penn State Worry Questionnaire

The PSWQ is a validated 16-item questionnaire with high internal consistency (Cronbach α=.93) used to assess worry [55]. Individuals are asked to rate statements on a scale ranging from 1 (*not at all typical of me*) to 5 (*very typical of me*) [55]. Example items include "My worries overwhelm me" and "When I am under pressure I worry a lot" [55]. Scores range from 16 to 80, with higher scores indicating a higher degree of worry.

Multidimensional Assessment of Interoceptive Awareness

The Multidimensional Assessment of Interoceptive Awareness (MAIA) is a 32-item questionnaire that assesses 8 domains of interoceptive awareness: noticing, not distracting, not worrying, attention regulation, emotional awareness, self-regulation, body listening, and trusting (Cronbach α=.66-.87) [56]. On a scale of 0 (*never*) to 5 (*always*), individuals are asked how often each statement applies to them generally in daily life [56]. Statements include "I distract myself from sensations of discomfort" and "I trust my body sensations" [56]. Scores range from 0 to 160, with higher scores indicating greater interoceptive awareness.

Safety and Adverse Events

Monitoring of safety occurred continuously during the study, and if an adverse event was reported, follow-up was conducted via phone by the project director using National Institute of Mental Health reportable events templates [57]. A final check was conducted by the project director at the conclusion of the study to assess adverse events potentially related to the intervention. If an event was reported, the same process was followed.

Sample Size

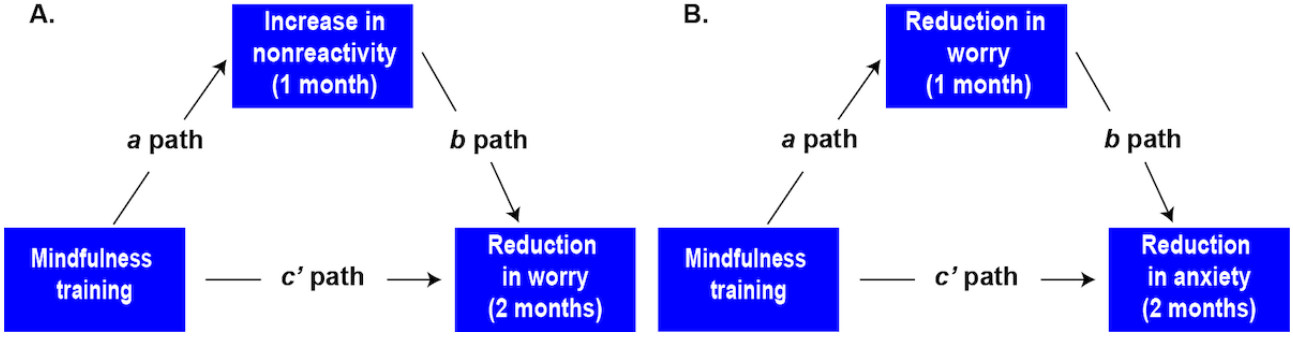
As there are no prior studies using the intervention, the target sample size of 65 was determined using pilot data. A pilot study of individuals with GAD-7 scores >9 (n=17) demonstrated a 39% reduction in the scores after completing 21 modules. Assuming a 10% reduction in the TAU group, a 1-tailed t test determined that a sample size of 52 would have 80% power with 1-sided 5% type I error to detect a statistically significant between-group difference in GAD-7 scores (Cohen d=0.7). We recruited 65 individuals to account for 25% attrition.

Statistical Analysis

Analyses were conducted using R (version 3.4.1), and a modified intent-to-treat approach was employed. This was defined as all participants who downloaded the app and completed all study assessments, regardless of treatment completion. A robust, mixed analysis of variance (ANOVA) was used to evaluate the primary outcomes, changes in anxiety, and emotional reactivity 2 months after treatment initiation in the TAU+app-delivered MT group relative to those in the TAU group. This and other robust statistical tests were chosen to avoid violating underlying model assumptions, such as normality. Because of the use of these nonparametric tests, the median and IQR were calculated. Post hoc comparisons of single effects were performed using the WRS2 package in R. In addition, Mann-Whitney U tests were performed, and the Hodges-Lehmann estimate of location shift was used to calculate the difference between groups at each time point [58]. Bonferroni correction was used to adjust for multiple comparisons. Effect sizes (r) were calculated by dividing the z score by the square root of the sample size using the Cohen criteria for r, where 0.1 is small, 0.3 is medium, and 0.5 is large [59].

We then conducted exploratory mediation analyses to evaluate (1) model A, whether increased nonreactivity mediated the relationship between MT and reduced worry, and (2) model B, whether reduced worry mediated the relationship between MT and reduced anxiety. To reduce the impact of the unit on the comparison of direct and indirect effects, the variables were standardized to have 0 means and unit SD. To have causal interpretations, the mediation models were built on the longitudinal [60]: In model A, MT was the independent variable, worry was the dependent variable, and nonreactivity was the mediating variable (Figure 1); in model B, MT was the independent variable, anxiety was dependent variable, and worry was the mediating variable (Figure 1). We calculated the direct and indirect effects, for which the SEs and 95% CI were computed using the bootstrap method with 1000 bootstrapped resamples [61].

Figure 1. (A) Path model for longitudinal causal mediation evaluating if increases in nonreactivity mediate the relationship between MT and reduction in worry. (B) Path model for longitudinal causal mediation evaluating if reduction in worry mediates the relationship between MT and reduction in anxiety. MT: mindfulness training.



The median and IQR were calculated to evaluate the engagement or the total number of modules completed. To explore the impact of anxiety on engagement at 2 months after treatment initiation, a robust regression model based on an M-estimator, which uses iteratively reweighted least squares estimation, was fitted, with the anxiety score as the independent variable and the total number of modules completed as the dependent variable.

Number Needed to Treat

The NNT, defined as the total number of individuals who need to receive treatment to prevent 1 adverse event, is a standard epidemiological measure used to communicate the effectiveness of a health care intervention [62]. The inverse of the absolute risk reduction was calculated by subtracting the total percentage of individuals who achieved remission (GAD-7 score ≤5, minimal anxiety) in the TAU group from the total percentage of those in the TAU+app-delivered MT group.

Reliable Change Index

Unlike statistical significance, clinical significance has traditionally lacked a consistent definition [63]. To address this need, a reliable change index (RCI) was created to evaluate the reliability of clinically significant changes [63]. We used the method developed by Jacobson and Truax to calculate the RCI for changes in anxiety scores [64]. If the RCI exceeded the *z*-scored level of significance from –1.96 to +1.96 ($P<.05$), we evaluated the percentage of participants with clinically significant change who met or exceeded it at 1 and 2 months after treatment initiation [63].

Risk of Bias

Six areas of potential bias across 7 domains were assessed using the Cochrane Collaboration tool for evaluating the risk of bias [65].

Results

Participants

We recruited 65 participants, obtained their consent to participate, and randomized them between May 2019 and October 2019. Baseline demographic characteristics are reported in Table 1. Of the 65 participants, 61 completed the study and were included in the modified intent-to-treat analysis (Figure 2). Before treatment initiation, 30% (19/63), 25% (16/63), and 32% (20/63) of participants reported comorbid anxiety, depression, and anxiety and depressive disorders, respectively (Table 1).

Table 1. Baseline demographic characteristics (N=63).

Characteristics	TAU[a]+app-delivered MT[b] participants (n=30)	TAU participants (n=33)
Age (years), mean (SD)	43 (15)	41 (16)
Sex, n (%)		
Male	2 (7)	3 (9)
Female	28 (93)	29 (88)
Other	0 (0)	1 (3)
Highest level of education completed, n (%)		
High school graduate or equivalent (eg, GED[c])	0 (0)	1 (3)
Some college or technical school	7 (23)	5 (15)
Associate degree	2 (7)	3 (9)
Bachelor's degree	7 (23)	16 (49)
Master's degree	13 (43)	8 (24)
Doctorate	1 (3)	0 (0)
Work status, n (%)		
Full-time	17 (57)	15 (46)
Part-time	3 (10)	9 (27)
Unemployed for <1 month	1 (3)	2 (6)
Unemployed for >1 month	3 (10)	3 (9)
Never employed	0 (0)	1 (3)
Not in labor force	6 (20)	3 (9)
Marital status, n (%)		
Never married	9 (30)	13 (39)
Married or cohabiting	16 (53)	18 (54)
Separated or divorced	4 (13)	2 (6)
Widowed	1 (3)	0 (0)
Race and ethnicity, n (%)		
White	27 (90)	28 (85)
Black	1 (3)	1 (3)
Asian	0 (0)	1 (3)
White, American Indian, or Alaskan native	1 (3)	0 (0)
White and Black	0 (0)	2 (6)
Hispanic, White, American Indian, or Alaskan native	1 (3)	1 (3)
Comorbid conditions, n (%)		
Anxiety disorder or disorders	10 (33)	9 (27)
Depressive disorder or disorders	10 (33)	6 (18)
Anxiety and depressive disorder or disorders	7 (23)	13 (39)
None	3 (10)	5 (15)
Concomitant medications, n (%)		
Selective serotonin reuptake inhibitors	6 (20)	3 (9)
SNRIs[d]	3 (3)	3 (9)
Other	2 (7)	6 (18)
>1 medication	2 (7)	5 (15)

Characteristics	TAU[a]+app-delivered MT[b] participants (n=30)	TAU participants (n=33)
None	17 (57)	17 (52)

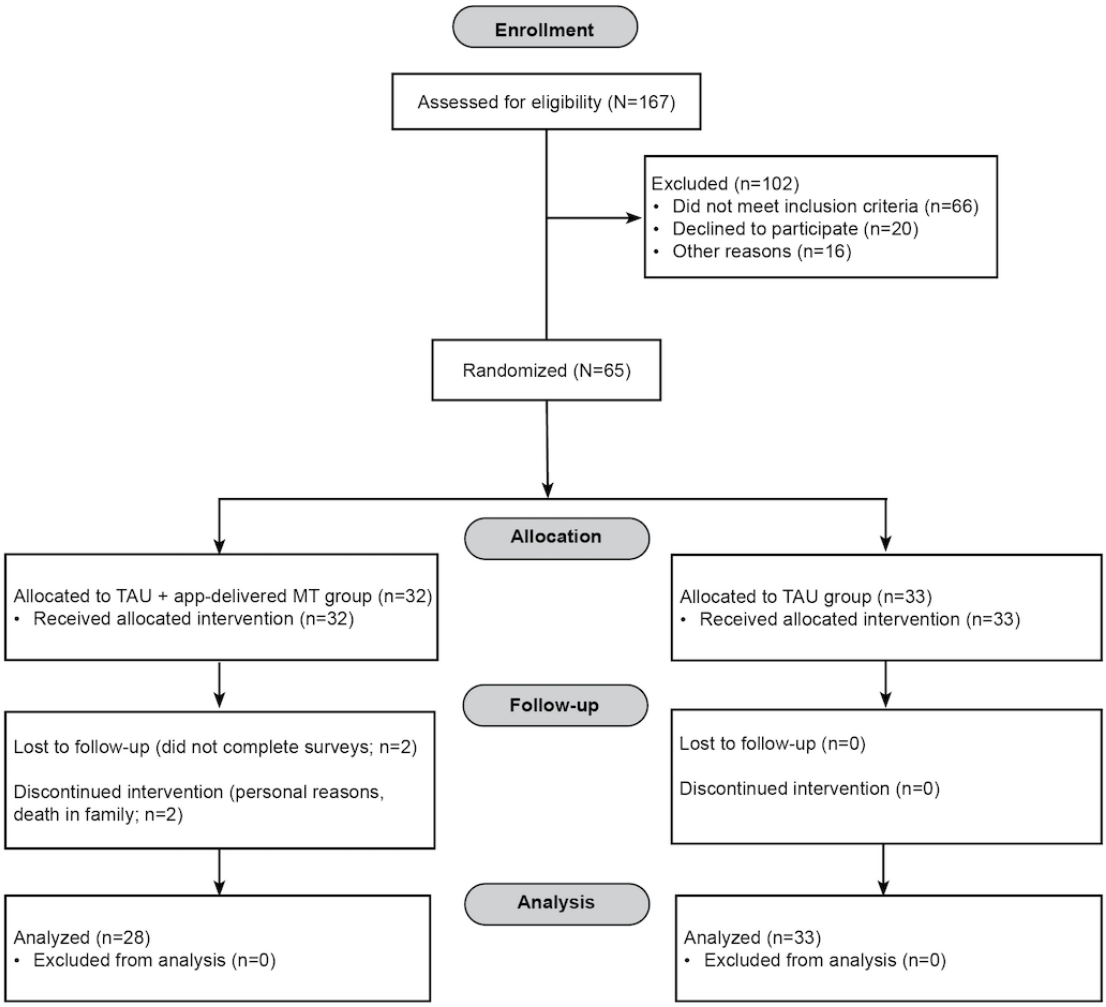
[a]TAU: treatment as usual.

[b]MT: mindfulness training.

[c]GED: general educational development.

[d]SNRIs: serotonin and norepinephrine reuptake inhibitors.

Figure 2. Participant flow diagram. MT: mindfulness training; TAU: treatment as usual.



Safety

There were no adverse events in the TAU group and 11% (3/28) of adverse events in the TAU+app-delivered MT group (2/3, 66% *anxiety* and 1/3, 33% *back pain*).

Changes in Outcome Measures

Effects of Intervention on Anxiety (GAD-7)

Baseline GAD-7 scores indicated moderate (36/63, 57%) to severe (24/63, 38%) anxiety in individuals with GAD (TAU+app-delivered MT: median 12, IQR 8; TAU: median 13, IQR 7). To examine the effect of MT on reduction in anxiety, we fitted a robust mixed ANOVA with group as the between-subjects factor, time as the within-subjects factor, and GAD-7 score as the dependent variable. We found a main effect of group ($F_{1,39.99}$=22.54; P<.001) and time ($F_{2,33.49}$=29.98; P<.001), with a significant group×time interaction ($F_{2,33.49}$=11.19; P<.001). At 1 month after treatment initiation, there was a significant difference between groups (P<.001; r=0.59) and the Hodges-Lehmann estimate, the nonparametric estimate of population change, was 5 (95% CI 4-7). The TAU+app-delivered MT group reported a median reduction in

anxiety scores of 5 (IQR 7.3; *P*<.001; *r*=0.89) compared with no change in the TAU group. At 1 month, the calculated RCI for the TAU+app-delivered MT group was –4.6 and a reliable change was seen in 64% (18/28) of the participants, while the calculated RCI for TAU was 0. A significant between-group difference (*P*<.001; *r*=0.68) was maintained at 2 months, and the Hodges-Lehmann estimate was 6 (95% CI 5-8; Figure 3); the TAU+app-delivered MT group reported a median reduction in anxiety scores of 8.5 (IQR 6.5; *P*<.001; *r*=0.96), while the TAU group reported a median reduction of 1 (IQR 5; *P*=.01; *r*=0.37), representing a 67% versus a 14% reduction. The RCI was –7.9 for the TAU+app-delivered MT group and reliable change was seen in 54% (15/28) of the participants. The RCI was –0.9 for the TAU group. See Table 2 for medians and IQRs, in addition to the means and SDs.

Figure 3. (A) Change in GAD-7 scores. (B) Change in nonreactivity scores. (C) Change in PSWQ scores. (D) Change in MAIA scores. GAD-7: Generalized Anxiety Disorder 7-item; MAIA: Multidimensional Assessment of Interoceptive Awareness; PSWQ: Penn State Worry Questionnaire.

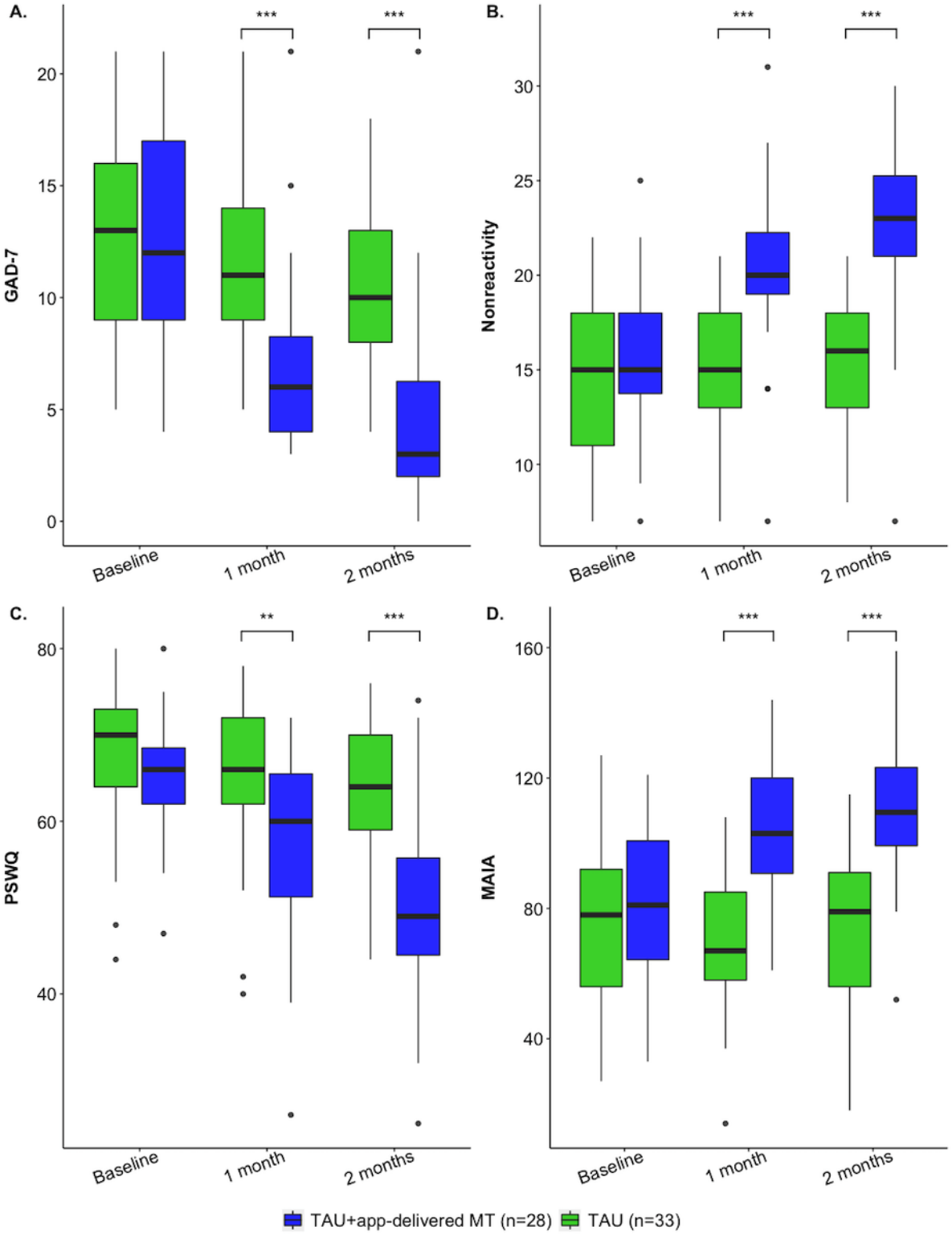


Table 2. Group-wise results for General Anxiety Disorder 7-item, nonreactivity subscale, Penn State Worry Questionnaire, and Multidimensional Assessment of Interoceptive Awareness (N=61).

Timepoints	TAU[a]+app-delivered MT[b] participants (n=28)		TAU participants (n=33)		*P* value[c]	Effect sizes (*r*)
	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)		
Generalized Anxiety Disorder-7						
Baseline	12.0 (8.0)	12.9 (4.8)	13.0 (7.0)	12.6 (4.3)	>.99	0
1 month	6.0 (4.3)	7.0 (4.1)	11.0 (5.0)	12.0 (3.7)	<.001	0.59
2 months	3.0 (4.3)	4.8 (4.1)	10.0 (5.0)	10.6 (3.5)	<.001	0.68
Δ[d] at 1 month (%)	–5.0 (–49)	–5.9 (–41)	0.0 (0)	–0.6 (3)	<.001	0.53
Δ at 2 months (%)	–8.5 (–67)	–8.1 (–60)	–1.0 (–14)	–2.0 (–10)	<.001	0.55
Nonreactivity						
Baseline	15.0 (4.3)	15.4 (4.3)	15.0 (7.0)	14.5 (4.5)	>.99	0
1 month	20.0 (3.3)	20.0 (4.3)	15.0 (5.0)	15.0 (4.0)	<.001	0.53
2 months	23.0 (4.3)	22.5 (4.8)	16.0 (5.0)	15.6 (3.4)	<.001	0.67
Δ at 1 month (%)	5.0 (36)	4.6 (35)	0.0 (0)	0.4 (7)	<.001	0.48
Δ at 2 months (%)	7.5 (51)	7.1 (52)	1.0 (8)	1.1 (15)	<.001	0.57
Penn State Worry Questionnaire						
Baseline	66.0 (6.5)	65.4 (7.0)	70.0 (9.0)	67.8 (8.0)	.26	0.14
1 month	60.0 (14.3)	57.8 (11.2)	66.0 (10.0)	65.5 (8.7)	<.001	0.33
2 months	49.0 (11.3)	49.9 (11.5)	64.0 (11.0)	63.8 (7.9)	<.001	0.55
Δ at 1 month (%)	–7.5 (–11)	–7.6 (–12)	–3.0 (–4)	–2.3 (–3)	.02	0.34
Δ at 2 months (%)	–15.0 (–23)	–15.5 (–23)	–3.0 (–5)	–4.0 (–6)	<.001	0.56
Multidimensional Assessment of Interoceptive Awareness						
Baseline	81.0 (36.5)	80.9 (23.2)	78.0 (36.0)	75.4 (26.0)	>.99	0
1 month	103.0 (29.3)	103.1 (21.9)	67.0 (27.0)	69.6 (21.2)	<.001	0.67
2 months	109.5 (24.0)	112.2 (22.8)	79.0 (35.0)	74.3 (23.4)	<.001	0.87
Δ at 1 month (%)	22.0 (25)	22.2 (39)	0.0 (0)	–5.8 (–4)	<.001	0.60
Δ at 2 months (%)	26.0 (29)	31.3 (53)	–2.0 (–2)	–1.1 (1)	<.001	0.85

[a]TAU: treatment as usual.

[b]MT: mindfulness training.

[c]Adjusted *P* values represent between-group comparisons.

[d]Δ: change between baseline and posttreatment.

Effects of Intervention on Nonreactivity (FFMQ Subscale)

To examine changes in nonreactivity, we performed a robust mixed ANOVA with group as the between-subjects factor, time as the within-subjects factor, and nonreactivity score as the dependent variable. This demonstrated a main effect of group ($F_{1,39.38}$=34.06; *P*<.001) and time ($F_{2,28.75}$=24.77; *P*<.001), with a significant group×time interaction ($F_{2,28.75}$=23.23; *P*<.001). At 1 month after treatment initiation, there was a significant difference between the groups (*P*<.001; *r*=0.53) and the Hodges-Lehmann estimate was –5 (95% CI –7 to –3). The TAU+app-delivered MT group reported a median increase of 5 (IQR 6.3) in nonreactivity scores (*P*<.001; *r*=0.95), whereas participants in the TAU group reported no change (Figure 3). A significant between-group difference (*P*<.001; *r*=0.67) was maintained at 2 months, and the Hodges-Lehmann estimate was –7 (95% CI –9 to –5); the TAU+app-delivered MT group reported a median increase of 7.5 (IQR 6) in nonreactivity scores (*P*<.001; *r*=0.95), while a median increase of 1 (IQR 6; *P*=.43, *r*=0.14) was seen in the TAU group.

Effects of Intervention on Worry (PSWQ)

To examine the effects of MT on worry, we ran a robust mixed ANOVA with group as the between-subjects factor, time as the within-subjects factor, and PSWQ score as the dependent variable. This revealed a main effect of group ($F_{1,37.85}$=19.66; *P*<.001) and time ($F_{2,27.12}$=34.78; *P*<.001), with a significant group×time interaction ($F_{2,27.12}$=10.30; *P*<.001). Participants in the TAU+app-delivered MT group reported a median

reduction in worry scores of 7.5 (IQR 8.5) at 1 month (P<.001; r=0.67; Figure 3), whereas the TAU group reported a median reduction of 3 (IQR 4; P=.01; r=0.44). There was a significant between-group difference (P<.001; r=0.55) at 2 months after treatment initiation, and the Hodges-Lehmann estimate was 14 (95% CI 9 to 19); the TAU+app-delivered MT group reported a median reduction in worry scores of 15 (IQR 14.3; P<.001; r=0.88) compared with a median reduction of 3 (IQR 6) reported by the TAU group (P<.001; r=0.61).

Effects of Intervention on Interoceptive Awareness (MAIA)

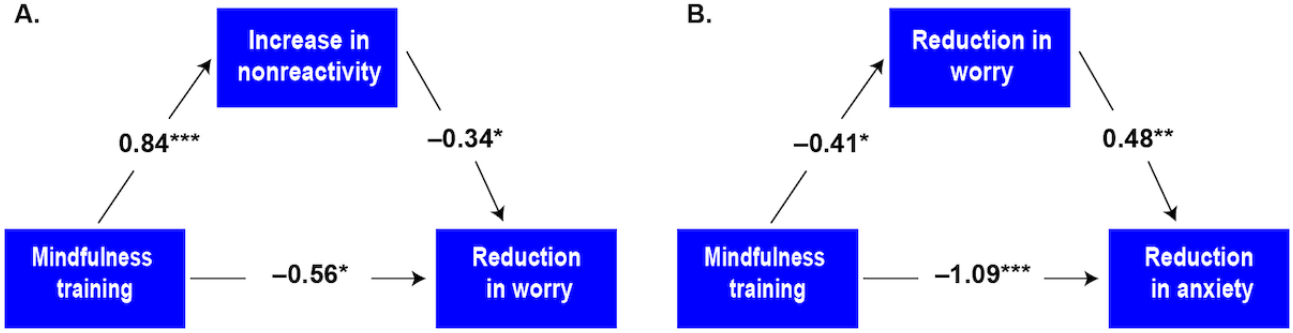
To examine changes in interoceptive awareness, we fitted a robust mixed ANOVA with group as the between-subjects factor, time as the within-subjects factor, and MAIA score as the dependent variable. We found a main effect of group ($F_{1,39.19}$=22.53; P<.001) and time ($F_{2,29.93}$=10.79; P<.001), with a significant group×time interaction ($F_{2,29.93}$=12.45; P<.001). At 1 month after treatment initiation, participants in the TAU+app-delivered MT group reported a median increase of 22 (IQR 30) in interoceptive awareness scores (P<.001; r=0.72), whereas the TAU group reported no change (median 0, IQR 18) in interoceptive awareness scores (P=.71; r=0.07; Figure 3). At 2 months, there was a significant between-group difference (P<.001; r=0.87), and the Hodges-Lehmann estimate was –13 (95% CI –15 to –10); the TAU+app-delivered MT group reported a median increase of 26 (IQR 28.5) in interoceptive awareness scores (P<.001; r=0.85), whereas the TAU group reported a median reduction of 2 (IQR 12; P>.99; r=0).

Mediation Analysis

Model A (Figure 4) shows the direct effect of MT on the reduction in worry and its indirect effect through nonreactivity. Reduction in worry and an increase in nonreactivity were defined as change in the PSWQ at 1 month and change in the nonreactivity scale from the FFMQ at 2 months after treatment initiation. Mediation analysis indicated that MT was related to a significant reduction in worry at 2 months with a direct effect of β=–.56 (SE=0.25; P=.03). MT also significantly increased nonreactivity (β=.84; SE=0.21; P<.001), which was significantly related to reduction in worry at 2 months (β=–.34; SE=0.14; P=.01). This implies that the relationship between MT and reduction in worry was partially mediated by an increase in nonreactivity ($\beta_{\text{indirect effect}}$=.84×–.34=–.29; 95% CI –0.68 to –0.04; P=.02). The total effect of MT on reduction in worry at 2 months was estimated to be β=–.85 (SE=0.23; P<.001). No effects were observed in the control group.

Model B (Figure 4) shows the relationship between MT and reduction in anxiety with a direct effect and an indirect effect mediated by a reduction in worry. These indicate that MT also had a significant impact on reduction in anxiety at 2 months, with a total effect of β=–1.28 (SE=0.26; P<.001). The direct effect of MT was estimated to be β=–1.09 (SE=0.25; P<.001). MT was related to significant reductions in worry at 1 month (β=–.41; SE=0.20; P=.04), whereas the latter was significantly related to reductions in anxiety at 2 months (β=.48; SE=0.16; P=.004). Thus, reductions in worry partially mediated the relationship between MT and reduction in anxiety ($\beta_{\text{indirect effect}}$=–.41×.48=–.19; 95% CI –0.40 to –0.02; P=.03). No effect was observed in the control group.

Figure 4. (A) Longitudinal causal mediation model with standardized regression coefficients illustrates how an increase in nonreactivity mediates the effects of mindfulness training on reduction in worry. (B) Longitudinal causal mediation model with standardized regression coefficients illustrates how a reduction in worry mediates the effects of mindfulness training on reduction in anxiety. *P<.05; **P<.01; ***P<.001.



Engagement

To evaluate program engagement, we calculated the median and IQR. At 1 month, the median number of modules completed was 18 (IQR 16.3). At 2 months, it was 25.5 (IQR 17), and 46% (13/28) participants completed the program. To explore the association between anxiety and module completion, we fitted a robust regression model and found that for each additional completed module demonstrating further progression through intervention, anxiety scores decreased by 1.37 (β=–1.37; SE=0.23; P<.001). The adjusted R^2 value for this model was 0.25.

Number Needed to Treat

At 2 months after treatment initiation, we found an NNT of 1.6: 64% (18/28) of the participants achieved remission in the TAU+app-delivered MT group compared with 3% (1/33) in the TAU group.

Risk of Bias

Using the Cochrane Collaboration criteria for evaluating bias [65], we found a low risk of bias in 6 of 7 categories, including random sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. Blinding of participants and personnel

was deemed to have a medium risk of bias because the project director and participants were unblinded to group allocation.

Discussion

Principal Findings

Anxiety is a debilitating and difficult-to-treat condition that affects hundreds of millions of people worldwide. Using a theory-based approach (targeting reinforcement learning), we developed a digital therapeutic that demonstrated a significant and clinically meaningful reduction in anxiety in individuals with GAD (NNT=1.6) [66]. This was confirmed by our finding that anxiety decreased with further progression through the program through the completion of more modules. Furthermore, we determined a potential mechanism underlying its effect: increases in mindfulness mediated decreases in worry and decreases in worry mediated reductions in anxiety. These effects were specific to the MT intervention. The mediation effect was significantly higher at 2 months (45%) than at 1 month (15%), which is consistent with participants having more exposure to the treatment (ie, a dose effect), a sleeper effect [67-69], or a combination of the two. These results are in direct alignment with the theoretical underpinnings that anxiety can be perpetuated through negative reinforcement—worry can feed back and perpetuate anxiety by introducing the *reward* of feeling more in control or temporarily distracting an individual from the aversive feelings of anxiety [14,15,21,42,70].

Treatment for Anxiety

Anxiety treatment has largely relied on antidepressant medications and psychotherapy (eg, CBT). These have yielded medium effect sizes for anxiety [9]. RCI is increasingly being used in treatment studies to assess whether changes are clinically significant [63]. A longitudinal study of low-intensity CBT using an RCI of ≤5 demonstrated a reliable change in 43.8% (181/439) of the participants [71]; our study found that 64% (18/28) and 54% (15/28) of the participants demonstrated reliable change at 1- and 2-months posttreatment using RCIs of 4.6% and 7.9%, respectively. For GAD, the NNT with antidepressants was 5.15. In this study, we found that specifically targeting a mechanistic pathway yielded large effect sizes with an NNT of 1.6. A previous single-arm pilot study in physicians with comparable levels of anxiety (median baseline GAD-7 score of 11.5) showed a similar magnitude of reduction in anxiety (57% reduction at 3 months) using the same app-delivered MT program [42]. This randomized controlled trial extends previous results and broadens these findings beyond anxious physicians to individuals with moderate to severe anxiety.

The Psychological Mechanisms of Anxiety

Potential mechanisms underlying anxiety have been hypothesized for over a century; yet, refinement in recent decades has opened the door for specific hypothesis testing [14,15]. For example, Mkrtchian et al [18] recently demonstrated avoidance as a part of reinforcement learning pathways in individuals with anxiety disorders. Our findings provide an important extension of these results by showing that targeting worry and avoidance yields clinically meaningful reductions in anxiety. These results and the finding that individuals reported increases in interoceptive awareness (measured by the MAIA) are in line with broader theoretical mindfulness frameworks that suggest that MT helps individuals learn to become more aware of and observe unpleasant emotions with awareness imbued with curiosity [72].

App-Delivered MT Targets Worry

MT may promote decentering, defined as a "metacognitive capacity to observe items that arise in the mind as mere psychological events" [73]. Decentering may help individuals disengage from perseverative worry habit loops that are perpetuated through reinforcement learning [46-48,74]. Our findings show that increases in mindfulness directly mediate the effects of app-delivered MT on reductions in worry. This may be the case possibly because of helping individuals step out of perseverative worry habit loops that are at the core of GAD and, in doing so, reduce their reinforcement. Furthermore, our results show that reductions in worry mediate the effects of MT on reductions in anxiety.

Practical Implications

The high prevalence of anxiety "vastly exceeds the capacities of mental health services," and this gap has only increased over the past several years [75]. App-based digital therapeutics offer a viable and practical route toward augmenting traditional mental health care and, in some cases, serve as a first-line treatment [76]. For example, if a patient in a primary care clinic screens positive for anxiety, an evidence-based digital therapeutic such as the one described in this study can be offered as an augmentation to standard medication treatment, or in some cases, it can be offered if a trial of medications has yielded suboptimal results or if a patient is not interested or willing to try a medication as an alternative. In addition, for individuals who are concerned about the confidentiality of mental health care (eg, feeling the need to ask a boss for regular time off for therapy visits), Health Insurance Portability and Accountability Act compliant digital therapeutics can offer discretion, privacy, and convenience. As integrative care models (eg, embedding psychiatric or psychological services within primary care clinics) gain momentum, one of the primary limitations is the cost and availability of trained therapists. However, because 85% of the US population has a smartphone, digital therapeutics may be able to serve as the *mobile* component of an integrative care clinic at a low cost, filling in for the lack of physical space and trained mental health clinicians [76]. In addition, in corporate settings where employers are increasingly aiming to meet the mental health needs of employees, it may be possible to quickly and confidentially deploy evidence-based digital therapeutics to help employees with mild or moderate anxiety. For employees with severe anxiety, who may have to wait several months to see their doctor or to obtain a mental health referral, a digital therapeutic may serve as a bridge to therapy or even a first-line treatment.

TAU Condition

We chose the GAD-7 as an outcome measure because it is widely used in clinical practice, yielding results that are interpretable in nonresearch settings. We chose TAU because

clinicians deliver standard treatment, such as prescribing a medication, and bolster these with an additional medication or recommendation for psychotherapy if a patient does not achieve a reduction in symptoms (ie, the TAU+X *add on* model). Although far from perfect as a control condition, TAU is standardly used in pragmatic clinical trials for these and other reasons [77,78].

Although the TAU group showed a significant decrease in anxiety symptoms (14%), there may be several reasons why TAU failed to show a greater reduction in symptomatology or achieve remission (3% vs 64%). These include a higher NNT for current medications and current models in which medical practices are designed more as a *sick care* model, in which acute, physically based issues are prioritized over mental health despite clear advantages of integrating mental health care into primary care settings [79,80]. This study demonstrated a clear proof-of-concept trial of a mindfulness-based digital therapeutic to deliver specific theory-driven and mechanistically based treatment in a clinically relevant setting. Furthermore, we aimed to recruit a real-world population by minimizing the exclusion criteria, such as comorbid disorders. In this study, most of the individuals (84%) presented with comorbid disorders, such as depression, which is consistent with how individuals present in primary care settings and to treatment specialists.

Limitations

This study has several strengths and notable findings, including designing for real-world applicability (eg, including individuals with comorbid disorders and concomitant medication treatment), accounting for engagement, assuring adequate sample sizes, registering outcomes, and minimizing the risk of bias. However, this study has some limitations. The TAU+ model was chosen to closely match the treatment a patient would encounter in a clinical setting. Whereas standard clinical care is highly variable, the study was designed such that randomization would ensure that this variability was equally distributed between the groups. Future studies using an active comparator (eg, CBT-based app) to control for attentional effects, longer follow-up periods, and incorporating multiple sites are needed to confirm the efficacy of this program. Second, this study sample comprised 90% (57/63) women. Although future sex-balanced studies are needed to determine the generalizability of these findings, women are twice as likely to develop an anxiety disorder and have a higher lifetime prevalence of GAD (7.1% vs 4.2%) than men [81,82]. Third, this study was designed to evaluate anxiety symptoms at 2 months after treatment initiation. Long-term follow-up studies are necessary to establish the long-term effects. Furthermore, although this study identifies potential psychological mechanisms of app-delivered MT, such as increased mindfulness mediating decreases in worry and anxiety, future studies are needed to explore its neurobiological mechanisms. In addition, studies performed in research laboratory settings (eg, National Institutes of Health stages I and II) may select more motivated individuals in general. Although randomization controls for between-group differences, real-world efficacy (National Institutes of Health stage III) is needed as the next step to determine efficacy in clinical settings.

Conclusions

In summary, for a large portion of the world's population that is affected by moderate to severe anxiety, targeted and mechanistically based treatments are needed. By combining theory and a new field of treatment delivery (digital therapeutics), we found that app-delivered MT significantly reduced anxiety, and its effects were mediated by increases in psychological nonreactivity and reductions in worry, suggesting a specific targeting of reinforcement learning.

Acknowledgments

This study was supported by the National Institute of Mental Health grant: R41MH118130.

Conflicts of Interest

JAB and AR are paid advisers to Sharecare, the company that owns the mindfulness app used in this study. This financial interest has been disclosed to and is being managed by Brown University, in accordance with its Conflict of Interest and Conflict of Commitment policies, including being restricted from recruitment, being blinded to the study group until after analysis, and not having access to data or performing analyses.

Multimedia Appendix 1
Unwinding Anxiety module outline.
[PDF File (Adobe PDF File), 137 KB - jmir_v23i12e26987_app1.pdf]

Multimedia Appendix 2
Overview of the app-delivered mindfulness training program, Unwinding Anxiety.
[PDF File (Adobe PDF File), 67784 KB - jmir_v23i12e26987_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1253 KB - jmir_v23i12e26987_app3.pdf]

References

1. Data Table 1: Lifetime prevalence of DSM-IV/WMH-CIDI disorders by sex and Cohort. Lifetime Prevalence of DSM-IV/WMH-CIDI Disorders by Sex and Cohort. 2017. URL: https://www.hcp.med.harvard.edu/ncs/ftpdir/NCS-R_Lifetime_Prevalence_Estimates.pdf [accessed 2021-10-25]
2. 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]
3. 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]
4. Asmundson GJ, Taylor S. How health anxiety influences responses to viral outbreaks like COVID-19: what all decision-makers, health authorities, and health care professionals need to know. J Anxiety Disord 2020 Apr;71:102211 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102211] [Medline: 32179380]
5. 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]
6. Generalised anxiety disorder and panic disorder in adults: management CG113. National Institute for Health and Care Excellence (NICE). 2011. URL: https://www.nice.org.uk/guidance/cg113 [accessed 2021-10-25]
7. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602 [FREE Full text] [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
8. Katzman MA. Current considerations in the treatment of generalized anxiety disorder. CNS Drugs 2009;23(2):103-120. [doi: 10.2165/00023210-200923020-00002] [Medline: 19173371]
9. Bandelow B, Reitt M, Röver C, Michaelis S, Görlich Y, Wedekind D. Efficacy of treatments for anxiety disorders: a meta-analysis. Int Clin Psychopharmacol 2015 Jul;30(4):183-192. [doi: 10.1097/YIC.0000000000000078] [Medline: 25932596]
10. Hoffman EJ, Mathew SJ. Anxiety disorders: a comprehensive review of pharmacotherapies. Mt Sinai J Med 2008 Jun;75(3):248-262. [doi: 10.1002/msj.20041] [Medline: 18704983]
11. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJH. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. World Psychiatry 2016 Oct;15(3):245-258 [FREE Full text] [doi: 10.1002/wps.20346] [Medline: 27717254]
12. Van Den Berg S, Shapiro DA, Bickerstaffe D, Cavanagh K. Computerized cognitive-behaviour therapy for anxiety and depression: a practical solution to the shortage of trained therapists. J Psychiatr Ment Health Nurs 2004 Oct;11(5):508-513. [doi: 10.1111/j.1365-2850.2004.00745.x] [Medline: 15450016]
13. Boeldt D, McMahon E, McFaul M, Greenleaf W. Using virtual reality exposure therapy to enhance treatment of anxiety disorders: identifying areas of clinical adoption and potential obstacles. Front Psychiatry 2019 Oct 25;10:773 [FREE Full text] [doi: 10.3389/fpsyt.2019.00773] [Medline: 31708821]
14. Borkovec TD, Roemer L. Perceived functions of worry among generalized anxiety disorder subjects: distraction from more emotionally distressing topics? J Behav Ther Exp Psychiatry 1995 Mar;26(1):25-30. [doi: 10.1016/0005-7916(94)00064-s] [Medline: 7642757]
15. Borkovec T, Ray W, Stöber J. Worry: a cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. Cognit Ther Res 1998;22(6):561-576. [doi: 10.1023/A:1018790003416]
16. Roemer L, Salters K, Raffa SD, Orsillo SM. Fear and avoidance of internal experiences in GAD: preliminary tests of a conceptual model. Cogn Ther Res 2005 Feb;29(1):71-88. [doi: 10.1007/s10608-005-1650-2]
17. Sibrava N, Borkovec T. The cognitive avoidance theory of worry. In: Davey G, Wells A, editors. Worry and its Psychological Disorders: Theory, Assessment and Treatment. Hoboken, New Jersey, United States: John Wiley & Sons Ltd; 2006:239-256.
18. Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling avoidance in mood and anxiety disorders using reinforcement learning. Biol Psychiatry 2017 Oct 01;82(7):532-539 [FREE Full text] [doi: 10.1016/j.biopsych.2017.01.017] [Medline: 28343697]
19. Diagnostic and Statistical Manual of Mental Disorders (5th Ed.). Arlington, VA: American Psychiatric Association; 2013.
20. Borkovec TD, Robinson E, Pruzinsky T, DePree JA. Preliminary exploration of worry: some characteristics and processes. Behav Res Ther 1983;21(1):9-16. [doi: 10.1016/0005-7967(83)90121-3] [Medline: 6830571]
21. Hong RY. Worry and rumination: differential associations with anxious and depressive symptoms and coping behavior. Behav Res Ther 2007 Feb;45(2):277-290. [doi: 10.1016/j.brat.2006.03.006] [Medline: 16635479]
22. Brewer JA, Roy A, Deluty A, Liu T, Hoge EA. Can mindfulness mechanistically target worry to improve sleep disturbances? Theory and study protocol for app-based anxiety program. Health Psychol 2020 Sep;39(9):776-784 [FREE Full text] [doi: 10.1037/hea0000874] [Medline: 32833479]
23. Beck AT, Emery G, Greenberg RL. Anxiety Disorders and Phobias: A Cognitive Perspective. New York: Basic Books; 1985.

24. Arnsten AFT. Stress signalling pathways that impair prefrontal cortex structure and function. Nat Rev Neurosci 2009 Jun;10(6):410-422 [FREE Full text] [doi: 10.1038/nrn2648] [Medline: 19455173]
25. Goyal M, Singh S, Sibinga EM, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]
26. Parmentier FB, García-Toro M, García-Campayo J, Yañez AM, Andrés P, Gili M. Mindfulness and symptoms of depression and anxiety in the general population: the mediating roles of worry, rumination, reappraisal and suppression. Front Psychol 2019 Mar 8;10:506 [FREE Full text] [doi: 10.3389/fpsyg.2019.00506] [Medline: 30906276]
27. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York: Bantam Dell; 1990.
28. Brewer JA, Pbert L. Mindfulness as an emerging treatment for smoking and other addictions? J Fam Med 2015;2(4):1035 [FREE Full text]
29. Brewer JA, Mallik S, Babuscio TA, Nich C, Johnson HE, Deleone CM, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. Drug Alcohol Depend 2011 Dec 01;119(1-2):72-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.05.027] [Medline: 21723049]
30. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. J Behav Med 2018 Dec;41(2):160-173 [FREE Full text] [doi: 10.1007/s10865-017-9884-5] [Medline: 28918456]
31. Janes AC, Datko M, Roy A, Barton B, Druker S, Neal C, et al. Quitting starts in the brain: a randomized controlled trial of app-based mindfulness shows decreases in neural responses to smoking cues that predict reductions in smoking. Neuropsychopharmacology 2019 Aug 30;44(9):1631-1638 [FREE Full text] [doi: 10.1038/s41386-019-0403-y] [Medline: 31039580]
32. Servaas MN, Riese H, Ormel J, Aleman A. The neural correlates of worry in association with individual differences in neuroticism. Hum Brain Mapp 2014 Sep;35(9):4303-4315 [FREE Full text] [doi: 10.1002/hbm.22476] [Medline: 24532549]
33. Brewer JA, Garrison KA, Whitfield-Gabrieli S. What about the "Self" is processed in the posterior cingulate cortex? Front Hum Neurosci 2013 Oct 02;7:647 [FREE Full text] [doi: 10.3389/fnhum.2013.00647] [Medline: 24106472]
34. Brewer JA, Worhunsky PD, Gray JR, Tang Y, Weber J, Kober H. Meditation experience is associated with differences in default mode network activity and connectivity. Proc Natl Acad Sci U S A 2011 Dec 13;108(50):20254-20259 [FREE Full text] [doi: 10.1073/pnas.1112029108] [Medline: 22114193]
35. Brewer JA, Garrison KA. The posterior cingulate cortex as a plausible mechanistic target of meditation: findings from neuroimaging. Ann N Y Acad Sci 2014 Jan;1307:19-27. [doi: 10.1111/nyas.12246] [Medline: 24033438]
36. Garrison KA, Scheinost D, Worhunsky PD, Elwafi HM, Thornhill TA, Thompson E, et al. Real-time fMRI links subjective experience with brain activity during focused attention. Neuroimage 2013 Nov 01;81:110-118 [FREE Full text] [doi: 10.1016/j.neuroimage.2013.05.030] [Medline: 23684866]
37. Garrison KA, Santoyo JF, Davis JH, Thornhill TA, Kerr CE, Brewer JA. Effortless awareness: using real time neurofeedback to investigate correlates of posterior cingulate cortex activity in meditators' self-report. Front Hum Neurosci 2013;7:440 [FREE Full text] [doi: 10.3389/fnhum.2013.00440] [Medline: 23964222]
38. Hölzel BK, Carmody J, Evans KC, Hoge EA, Dusek JA, Morgan L, et al. Stress reduction correlates with structural changes in the amygdala. Soc Cogn Affect Neurosci 2010 Mar;5(1):11-17 [FREE Full text] [doi: 10.1093/scan/nsp034] [Medline: 19776221]
39. Crane RS, Kuyken W, Williams JM, Hastings RP, Cooper L, Fennell MJ. Competence in teaching mindfulness-based courses: concepts, development and assessment. Mindfulness (N Y) 2012 Mar 22;3(1):76-84 [FREE Full text] [doi: 10.1007/s12671-011-0073-2] [Medline: 23293683]
40. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
41. Carl JR, Miller CB, Henry AL, Davis ML, Stott R, Smits JA, et al. Efficacy of digital cognitive behavioral therapy for moderate-to-severe symptoms of generalized anxiety disorder: a randomized controlled trial. Depress Anxiety 2020 Dec 29;37(12):1168-1178. [doi: 10.1002/da.23079] [Medline: 32725848]
42. Roy A, Druker S, Hoge EA, Brewer JA. Physician anxiety and burnout: symptom correlates and a prospective pilot study of app-delivered mindfulness training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
43. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
44. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-34. [Medline: 9881538]
45. Qualtrics XM. 2021. URL: https://www.qualtrics.com/ [accessed 2021-10-25]

46. Brewer JA. Mindfulness training for addictions: has neuroscience revealed a brain hack by which awareness subverts the addictive process? Curr Opin Psychol 2019 Aug;28:198-203. [doi: 10.1016/j.copsyc.2019.01.014] [Medline: 30785066]
47. Brewer JA, Elwafi HM, Davis JH. Craving to quit: psychological models and neurobiological mechanisms of mindfulness training as treatment for addictions. Psychol Addict Behav 2013 Jun;27(2):366-379 [FREE Full text] [doi: 10.1037/a0028490] [Medline: 22642859]
48. Brewer JA, Ruf A, Beccia AL, Essien GI, Finn LM, van Lutterveld R, et al. Can mindfulness address maladaptive eating behaviors? Why traditional diet plans fail and how new mechanistic insights may lead to novel interventions. Front Psychol 2018;9:1418. [doi: 10.3389/fpsyg.2018.01418] [Medline: 30250438]
49. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2018 Jun 18:324-331. [doi: 10.1093/ntr/nty126] [Medline: 29917096]
50. 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]
51. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
52. Ruiz MA, Zamorano E, García-Campayo J, Pardo A, Freire O, Rejas J. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. J Affect Disord 2011 Feb;128(3):277-286. [doi: 10.1016/j.jad.2010.07.010] [Medline: 20692043]
53. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
54. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
55. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
56. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The Multidimensional Assessment of Interoceptive Awareness (MAIA). PLoS One 2012;7(11):e48230 [FREE Full text] [doi: 10.1371/journal.pone.0048230] [Medline: 23133619]
57. NIMH Clinical Research Toolbox. URL: https://www.nimh.nih.gov/funding/clinical-research/clinical-research-toolbox/nimh-clinical-research-toolbox [accessed 2021-10-25]
58. Hodges JL, Lehmann EL. Estimates of location based on rank tests. Ann Math Statist 1963 Jun;34(2):598-611. [doi: 10.1214/aoms/1177704172]
59. Field A, Miles J, Field Z. Discovering Statistics Using R. Thousand Oaks, CA: SAGE Publications Ltd; 2012:1-992.
60. Jose PE. The merits of using longitudinal mediation. Educ Psychol 2016 Sep 02;51(3-4):331-341. [doi: 10.1080/00461520.2016.1207175]
61. Imai K, Keele L, Tingley D, Yamamoto T. Causal mediation analysis using R. In: Vinod H, editor. Advances in Social Science Research Using R. New York: Springer; 2010:129-154.
62. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. Br Med J 1995 Feb 18;310(6977):452-454 [FREE Full text] [doi: 10.1136/bmj.310.6977.452] [Medline: 7873954]
63. Wise EA. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. J Pers Assess 2004 Feb;82(1):50-59. [doi: 10.1207/s15327752jpa8201_10] [Medline: 14979834]
64. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
65. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
66. Skinner BF. Operant behavior. Am Psychol 1963 Aug;18(8):503-515. [doi: 10.1037/h0045185]
67. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. Drug Alcohol Depend 2009 Feb 1;100(1-2):178-181 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.09.015] [Medline: 19041197]
68. Carroll KM, Easton CJ, Nich C, Hunkele KA, Neavins TM, Sinha R, et al. The use of contingency management and motivational/skills-building therapy to treat young adults with marijuana dependence. J Consul Clin Psychol 2006 Oct;74(5):955-966. [doi: 10.1037/0022-006x.74.5.955]
69. Epstein DH, Hawkins WE, Covi L, Umbricht A, Preston KL. Cognitive-behavioral therapy plus contingency management for cocaine use: findings during treatment and across 12-month follow-up. Psychol Addict Behav 2003 Mar;17(1):73-82 [FREE Full text] [doi: 10.1037/0893-164x.17.1.73] [Medline: 12665084]

70. Brewer JA, Roy A. Can approaching anxiety like a habit lead to novel treatments? Am J Lifestyle Med 2021 Apr 27;15(5):489-494. [doi: 10.1177/15598276211008144] [Medline: 34646095]
71. Ali S, Rhodes L, Moreea O, McMillan D, Gilbody S, Leach C, et al. How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. Behav Res Ther 2017 Jul;94:1-8. [doi: 10.1016/j.brat.2017.04.006] [Medline: 28437680]
72. Desrosiers A, Vine V, Klemanski DH, Nolen-Hoeksema S. Mindfulness and emotion regulation in depression and anxiety: common and distinct mechanisms of action. Depress Anxiety 2013 Jul 16;30(7):654-661 [FREE Full text] [doi: 10.1002/da.22124] [Medline: 23592556]
73. Fresco DM, Moore MT, van Dulmen MH, Segal ZV, Ma SH, Teasdale JD, et al. Initial psychometric properties of the experiences questionnaire: validation of a self-report measure of decentering. Behav Ther 2007 Sep;38(3):234-246. [doi: 10.1016/j.beth.2006.08.003] [Medline: 17697849]
74. Hoge EA, Bui E, Goetter E, Robinaugh DJ, Ojserkis RA, Fresco DM, et al. Change in decentering mediates improvement in anxiety in mindfulness-based stress reduction for generalized anxiety disorder. Cognit Ther Res 2015 Apr;39(2):228-235 [FREE Full text] [doi: 10.1007/s10608-014-9646-4] [Medline: 28316355]
75. Firth J, Torous J, Carney R, Newby J, Cosco TD, Christensen H, et al. Digital technologies in the treatment of anxiety: recent innovations and future directions. Curr Psychiatry Rep 2018 May 19;20(6):44. [doi: 10.1007/s11920-018-0910-2] [Medline: 29779065]
76. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
77. Löfholm CA, Brännström L, Olsson M, Hansson K. Treatment-as-usual in effectiveness studies: what is it and does it matter? Int J Soc Welf 2012 Feb 28;22(1):25-34. [doi: 10.1111/j.1468-2397.2012.00870.x]
78. Kazdin AE. Treatment as usual and routine care in research and clinical practice. Clin Psychol Rev 2015 Dec;42:168-178. [doi: 10.1016/j.cpr.2015.08.006] [Medline: 26431668]
79. Unützer J, Schoenbaum M, Druss BG, Katon WJ. Transforming mental health care at the interface with general medicine: report for the presidents commission. Psychiatr Serv 2006 Jan;57(1):37-47. [doi: 10.1176/appi.ps.57.1.37] [Medline: 16399961]
80. Thielke S, Vannoy S, Unützer J. Integrating mental health and primary care. Prim Care 2007 Sep;34(3):571-92, vii. [doi: 10.1016/j.pop.2007.05.007] [Medline: 17868760]
81. Catuzzi JE, Beck KD. Anxiety vulnerability in women: a two-hit hypothesis. Exp Neurol 2014 Sep;259:75-80. [doi: 10.1016/j.expneurol.2014.01.023] [Medline: 24518489]
82. Christiansen D. Examining sex and gender differences in anxiety disorders. In: Durbano F, editor. A Fresh Look at Anxiety Disorders. London, U.K: IntechOpen Limited; 2015.

Abbreviations

ANOVA: analysis of variance
CBT: cognitive behavioral therapy
FFMQ: Five Facet Mindfulness Questionnaire
GAD: generalized anxiety disorder
MAIA: Multidimensional Assessment of Interoceptive Awareness
MT: mindfulness training
NNT: number needed to treat
PSWQ: Penn State Worry Questionnaire
RCI: reliable change index
TAU: treatment as usual

Edited by R Kukafka, G Eysenbach; submitted 08.01.21; peer-reviewed by E Lattie, L Russell, SL Tan; comments to author 22.03.21; revised version received 06.04.21; accepted 29.09.21; published 02.12.21.

Please cite as:
Roy A, Hoge EA, Abrante P, Druker S, Liu T, Brewer JA
Clinical Efficacy and Psychological Mechanisms of an App-Based Digital Therapeutic for Generalized Anxiety Disorder: Randomized Controlled Trial
J Med Internet Res 2021;23(12):e26987
URL: https://www.jmir.org/2021/12/e26987
doi:10.2196/26987
PMID:34860673

Original Paper

A Theory- and Evidence-Based Digital Intervention Tool for Weight Loss Maintenance (NoHoW Toolkit): Systematic Development and Refinement Study

Marta M Marques[1,2], PhD; Marcela Matos[3*], PhD; Elina Mattila[4*], PhD; Jorge Encantado[1], PhD; Cristiana Duarte[5], PhD; Pedro J Teixeira[1], PhD; R James Stubbs[5], PhD; Falko F Sniehotta[6,7,8], PhD; Miikka Ermes[4], PhD; Marja Harjumaa[4], PhD; Juha Leppänen[4], PhD; Pasi Välkkynen[4], PhD; Marlene N Silva[9], PhD; Cláudia Ferreira[3], PhD; Sérgio Carvalho[3,10], PhD; Lara Palmeira[3], PhD; Graham Horgan[11], PhD; Berit Lilienthal Heitmann[12,13], PhD; Elizabeth H Evans[14], PhD; António L Palmeira[9], PhD

[1]CIPER -Interdisciplinary Centre for the Study of Human Performance, Faculty of Human Kinetics, University of Lisbon, Lisbon, Portugal

[2]Comprehensive Health Research Centre, NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal

[3]Center for Research in Neuropsychology and Cognitive and Behavioral Intervention (CINEICC), Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal

[4]VTT Technical Research Centre of Finland Ltd, Espoo, Finland

[5]School of Education, Language and Psychology, York St John University, York, United Kingdom

[6]Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, United Kingdom

[7]NIHR Policy Research Unit Behavioural Science, Newcastle University, Newcastle Upon Tyne, United Kingdom

[8]Faculty of Behavioural, Management and Social Sciences, University of Twente, Twente, Netherlands

[9]Centro de Investigação em Desporto, Educação Física e Saúde, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal

[10]HEI-Lab: Digital Human-Environment Interaction Lab, School of Psychology and Life Sciences, Lusófona University, Lisbon, Portugal

[11]Biomathematics and Statistics Scotland, Aberdeen, United Kingdom

[12]Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, The Capital Region, Denmark

[13]Department of Public Health, Section for General Practice, University of Copenhagen, Copenhagen, Denmark

[14]Department of Psychology, Durham University, Durham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Marta M Marques, PhD
Comprehensive Health Research Centre
NOVA Medical School
Universidade Nova de Lisboa
Rua do Instituto Bacteriológico, nº 5 l
Lisbon, 1150-190
Portugal
Phone: 351 218803105
Email: mmarques@nms.unl.pt

Abstract

Background: Many weight loss programs show short-term effectiveness, but subsequent weight loss maintenance is difficult to achieve. Digital technologies offer a promising means of delivering behavior change approaches at low costs and on a wide scale. The Navigating to a Healthy Weight (NoHoW) project, which was funded by the European Union's Horizon 2020 research and innovation program, aimed to develop, test, and evaluate a digital toolkit designed to promote successful long-term weight management. The toolkit was tested in an 18-month, large-scale, international, 2×2 factorial (motivation and self-regulation vs emotion regulation) randomized controlled trial that was conducted on adults with overweight or obesity who lost ≥5% of their body weight in the preceding 12 months before enrollment into the intervention.

Objective: This paper aims to describe the development of the NoHoW Toolkit, focusing on the logic models, content, and specifications, as well as the results from user testing.

Methods: The toolkit was developed by using a systematic approach, which included the development of the theory-based logic models, the selection of behavior change techniques, the translation of these techniques into a web-based app (NoHoW Toolkit components), technical development, and the user evaluation and refinement of the toolkit.

Results: The toolkit included a set of web-based tools and inputs from digital tracking devices (smart scales and activity trackers) and modules that targeted weight, physical activity, and dietary behaviors. The final toolkit comprised 34 sessions that were distributed through 15 modules and provided active content over a 4-month period. The motivation and self-regulation arm consisted of 8 modules (17 sessions), the emotion regulation arm was presented with 7 modules (17 sessions), and the combined arm received the full toolkit (15 modules; 34 sessions). The sessions included a range of implementations, such as videos, testimonies, and questionnaires. Furthermore, the toolkit contained 5 specific data tiles for monitoring weight, steps, healthy eating, mood, and sleep.

Conclusions: A systematic approach to the development of digital solutions based on theory, evidence, and user testing may significantly contribute to the advancement of the science of behavior change and improve current solutions for sustained weight management. Testing the toolkit by using a 2×2 design provided a unique opportunity to examine the effect of motivation and self-regulation and emotion regulation separately, as well as the effect of their interaction in weight loss maintenance.

(J Med Internet Res 2021;23(12):e25305) doi:10.2196/25305

KEYWORDS

mHealth; behavior change techniques; weight management; motivation; self-regulation; emotion regulation; self-monitoring; user testing; logic models

Introduction

Background

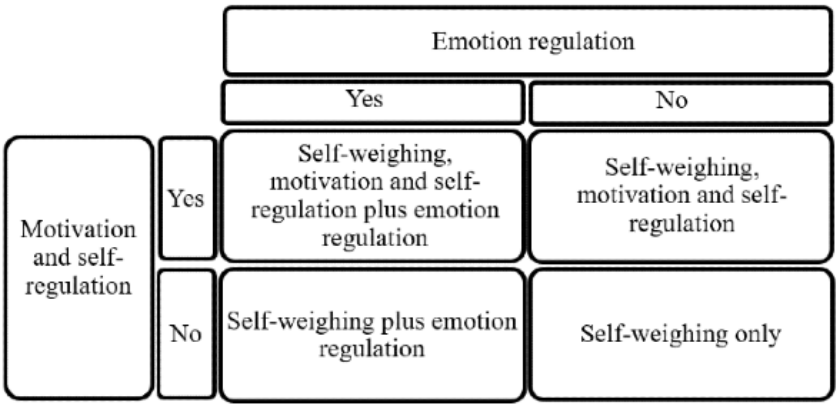
Behavior change interventions for overweight and obesity that target dietary behaviors, physical activity, and other weight management components show beneficial effects in reducing weight and improving health, at least in the short term [1,2], but most individuals experience weight regain over subsequent months and years [3,4]. Thus, a key challenge for such interventions is to find sustainable and scalable methods for long-term weight loss maintenance (WLM) [5,6]. The emergence and rapid growth of digital behavior change interventions (ie, sets of activities or products designed to change specified behavior patterns through digital technology) are a response to the urgent need for scalability and sustainability and can help to better understand the complexity behind individuals' decisions and engagement in behaviors that affect their health and well-being, including sustained weight management [1,7].

Although research on the effectiveness of digital behavior change interventions is at an early stage, existing reviews have found small effects and between-study variability on outcomes and behaviors [8,9]. These interventions can be optimized by identifying the features that contributed to their effectiveness through systematic theory- and evidence-based processes of intervention development, implementation, testing, and reporting (eg, Medical Research Council [10] and Template for Intervention Description and Replication guidance [11]).

In the context of weight management, the core features of effective interventions include techniques in line with self-regulation theories, such as goal-setting, self-monitoring of weight and behavior, feedback on behavior and weight, and plans to cope with risk factors for weight regain and relapse prevention (eg, problem-solving) [12-15]. Autonomous motivation has also been associated with change in energy balance behaviors for obesity management [16]. As the number of digital interventions based on motivation and self-regulation approaches to sustained weight management is limited [17], it is important to test if these theoretical approaches produce sustained behavior changes in the context of WLM. In addition, contextual-based emotion regulation strategies that promote compassion, acceptance, and mindfulness [18-20] may also have an impact on behavior changes that promote WLM [21]. With these considerations in mind, we developed a theory-based toolkit, delivered as a portfolio of embedded digital tracking technologies (smart scale and activity trackers), mini apps, and web resources accessible through a computer, tablet, and mobile phone, as part of the European Union's Horizon 2020 project *NoHoW: Evidence-Based Tools for Weight Loss Maintenance*.

The toolkit was used in the international Navigating to a Healthy Weight (NoHoW) innovative, 2×2 factorial, single-blind randomized controlled trial (RCT) for WLM. All 4 arms of the intervention included self-weighing and self-monitoring of activity using commercial Wi-Fi scales and activity trackers. The active control arm included access to generic weight management content, and the 3 intervention arms consisted of *motivation and self-regulation* content, *emotion regulation* content, and *combined motivation and self-regulation+emotion regulation* content (Figure 1; detailed information about the NoHoW project and trial can be found elsewhere [22]).

Figure 1. The Navigating to a Healthy Weight trial design.



Aim

This paper aims to describe the processes involved in the development of the NoHoW Toolkit to a stage appropriate for an RCT evaluation (trial registration number: ISRCTN88405328) and provide a detailed overview of the NoHoW Toolkit.

Methods

Systematic and Iterative Process

The development of the NoHoW Toolkit involved a multidisciplinary team of researchers in behavioral science, obesity and weight management, exercise and nutrition, software development, data analytics, biomathematics, and user experience (UX). It followed a systematic and iterative process incorporating both theory- and evidence-based approaches, as well as user testing to refine usability, and it also followed best practices by conducting the steps that are commonly recommended by the Behaviour Change Wheel [23], the intervention mapping approach [24], and the Medical Research Council [25] guidance for intervention development. It consisted of the following five steps: (1) development of theory-based logic models; (2) selection of the content, that is, behavior change techniques; (3) translation of these techniques into a coherent web-based app (NoHoW Toolkit components); (4) technical development; and (5) user evaluation and refinement of the toolkit for subsequent evaluation in the context of an RCT.

Step 1: Development of Theory-Based Logic Models

The selection of the psychological and behavioral factors to be targeted by the toolkit was based on existing studies on the most relevant theoretical frameworks in the context of physical activity, healthy eating, and weight management [2]. These studies were reviewed by the core research team and informed the NoHoW theoretical logic models that were developed. These logic models schematically represent the relationships among (1) the primary and secondary outcomes, including behavioral outcomes; (2) the theoretical mediators that were hypothesized to explain the effect of the toolkit content on the primary and secondary outcomes; (3) the content of the toolkit; and (4) the hypothesized moderators of the intervention effects. This task was executed by the behavioral science team with feedback from the members of the larger project consortium and the external advisory board.

Step 2: Selection of Behavior Change Techniques

The next developmental task was the identification and selection of the specific techniques that would form the content of the NoHoW Toolkit. This selection was derived from the studies reviewed (step 1) describing the behavior, motivation, and emotion regulation techniques hypothesized to have an impact on the theory-based mediators of WLM as represented in the logic models. For example, providing choice is a core technique to foster autonomy. We conducted an additional scoping review to identify the frequently used and effective intervention techniques and modes of delivery used in digital behavior change interventions for changing health behaviors in the context of long-term weight management (Multimedia Appendix 1). The selection of the final set of techniques was carried out through discussions among the team members. To ensure standardization in the description of the NoHoW Toolkit, the selected techniques were reported using the identifiers and labels from reliable taxonomies (when available) such as the Behavior Change Techniques Taxonomy (BCTT) version 1 (eg, BCTT 1.2 Problem Solving) [26] or the classification of Motivation and Behavior Change Techniques (MBCTs; eg, MBCT 6: Providing Choice) [27]. No taxonomy is yet available for the emotion regulation techniques.

Step 3: Translation of the Techniques Into the NoHoW Toolkit Components

Table 1 presents an overview of the development tasks (steps 3 to 5). Initially, the research team created user cases involving personas based on recent research identifying individuals involved in WLM [28] (who varied in age, gender, weight history, digital literacy, and reasons for participating in a WLM intervention) and scenarios (eg, first visit, throughout the intervention, and during a relapse) to describe the target users and their potential experiences when using the toolkit. This exercise provided the starting point for specifying the functional requirements of the toolkit and identifying how the techniques selected in step 3 would be implemented and also facilitated the development of a common language between the content development and technology development partners. A guidance manual was created to aid the teams involved in the development of the modular content of the toolkit. The manual stipulated that the content of each module should identify (1) specific behavior

goals; (2) theoretical constructs targeted; (3) techniques targeted; (4) rationale: how the goal, theoretical construct, and techniques are linked; and (5) the mode of delivery of each technique (eg, video or text) described in a way that would inform their implementation by the software developers and designers. Scripts were developed for each implementation.

Table 1. NoHoW Toolkit[a] development process.

Tasks	Team responsible
Task 1: persona and scenario development	Behavior change team and UX[b] team
Task 2: development of the overall content of the toolkit (eg, sessions) based on the motivation, self-regulation, and emotion regulation theories of behavior change	Behavior change team
Task 3: system architecture design	Software development team
Task 4: full description of each implementation (technique or clusters of techniques) for each content feature (eg, session)	Behavior change team
Task 5: development of a list of possible modalities in which the techniques would be implemented (eg, quiz, animation video, audio, text, and testimonial)	Behavior change team
Task 6: feedback about feasibility of implementation, engagement, and other options and details of the technical implementation	Software development team
Task 7: adjustments in accordance	Behavior change team
Task 8: UI[c] design	UX team and software development team
Task 9: functional description	Software development team
Task 10: programming	Software development team
Task 11: feasibility study plan	UX team
Task 12: toolkit sessions upload using a content management system	Behavior change team
Task 13: the implementation is user tested	Behavior change team and software development team
Task 14: final adjustments	Behavior change team and software development team

[a]NoHoW: Navigating to a Healthy Weight.

[b]UX: user experience.

[c]UI: user interface.

Step 4: Toolkit Technical Development

We identified that the 2 most important functionalities of the toolkit were (1) providing the WLM intervention in 3 theoretically informed, evidence-based versions and a control version and (2) supporting the digital tracking of weight, physical activity, and sleep. Given that the toolkit was intended to be used for several months, a positive user interaction and an attractive user interface (UI) were considered necessary. A UX designer was involved in the iterative development of the most crucial UI elements. The design process started with scenario development and identification of the user requirements for the toolkit based on the scenarios, informing the development of the initial UI concepts. The concepts were visualized by the UX designer and tested in small-scale interviews with participants who were similar to the target users. The UI for the intervention sessions was designed through collaboration between the behavior change team and the intervention designers. An admin UI was developed to serve the needs of the intervention designers and trial managers. The intervention designers needed to be able to add, refine, and update content continuously during the toolkit development as well as to define the content, rules, and schedules for different email prompts and reminders to be sent to the toolkit users. The trial managers needed to create new users of the toolkit, assign them to different intervention arms, and manage their status depending on whether they were still involved in the study.

Step 5: User Evaluation and Toolkit Refinement

We conducted 2 user evaluation studies. The first, which sought to identify key technical and UX-related issues and results from this study, informed the refinements to the NoHoW Toolkit version 1.0. This study was conducted in the United Kingdom with English-speaking adults, recruited primarily by advertisement or invitation from the University of Derby. The inclusion criteria were as follows: aged ≥18 years, able to travel to the University of Derby, able to follow written and verbal information in English, ability to access the internet, and currently overweight or have been overweight in the past, with at least one weight loss attempt. Approval was obtained from the research ethics committee of the University of Derby. A mixed methods approach was used for the analysis of the data through questionnaires, interviews, toolkit use log data, and data from wireless scales and activity trackers (Fitbit [Fitbit LLC]; see Multimedia Appendix 2 for the study description).

After refinement of the toolkit version 1.0 into version 2.0 and translation from English into Portuguese and Danish (languages of the 3 countries where the NoHoW trial was conducted), we conducted a second usability testing study with the Portuguese sample. This was a qualitative study consisting of in-depth

interviews, with the aim of understanding users' motivation to engage with the toolkit and determining which functionalities of the toolkit could promote user engagement (see Multimedia Appendix 3 for the full description). In these interviews, conducted after changes were implemented on top of toolkit version 1.0, the participants were shown a beta version of the toolkit, which included a demonstration of the navigation procedures, and 1 session of the toolkit. The interviews, which lasted approximately 1 hour, were audio-recorded and transcribed verbatim. Thematic analysis was conducted using MAXQDA version 12 (VERBI GmbH). Approval was obtained from the research ethics committee of the University of Lisbon.

Results

Step 1: Development of Theory-Based Logic Models

Available evidence from behavioral weight management interventions [17] showed some support for the effectiveness of interventions based on self-regulation theories [13,14,29] and self-determination theory [16]. There is also increasing evidence showing the impact of emotions in weight management [30-34] and indicating that contextual behavioral approaches can contribute to psychological well-being and weight management through the development of skills aimed at reducing the automaticity of maladaptive patterns of behavior (eg, mindfulness, acceptance, and compassion toward internal negative states) [34-38].

The available evidence formed the basis for the NoHoW research hypotheses that WLM could be supported by strategies that promote (1) self-regulation (setting optimal goals and reviewing them, action and coping planning, and action control) and motivation factors (promotion of autonomous motivation vs controlled motivation, intrinsic goals, and flexible regulation to change behaviors and maintain weight loss), (2) emotion regulation factors that may undermine self-management of energy balance–related behaviors (reduce weight-related shame and self-criticism, reduce difficulties in emotion regulation, and increase psychological flexibility, mindfulness, and compassion) and (3) interactions between (1) and (2).

The NoHoW logic models, presented in Figures 2-4, schematically represent the relationships derived from these theoretical approaches. We developed 3 logic models, one for each of the intervention arms: *motivation and self-regulation* arm, *emotion regulation* arm, and the *combined* arm.

Figure 2. Logic model: motivation and self-regulation arm.

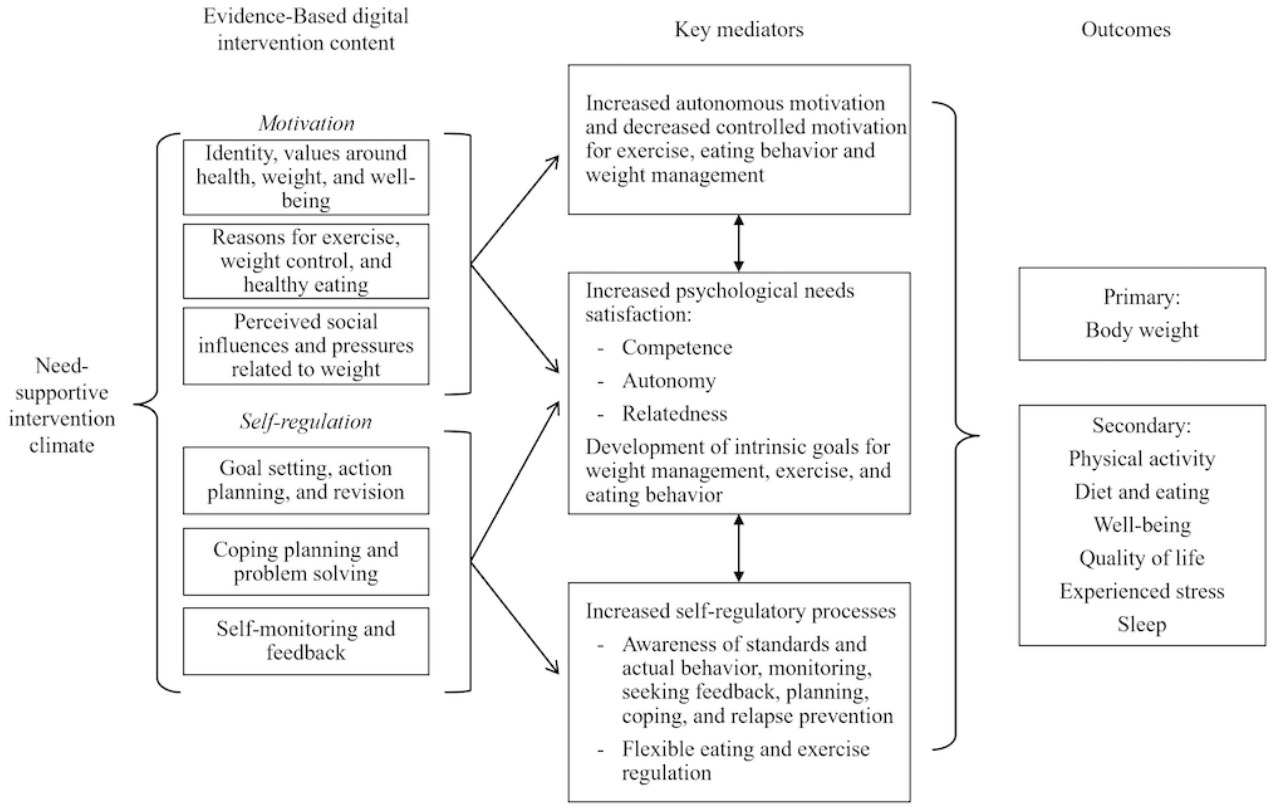


Figure 3. Logic model: emotion regulation arm.

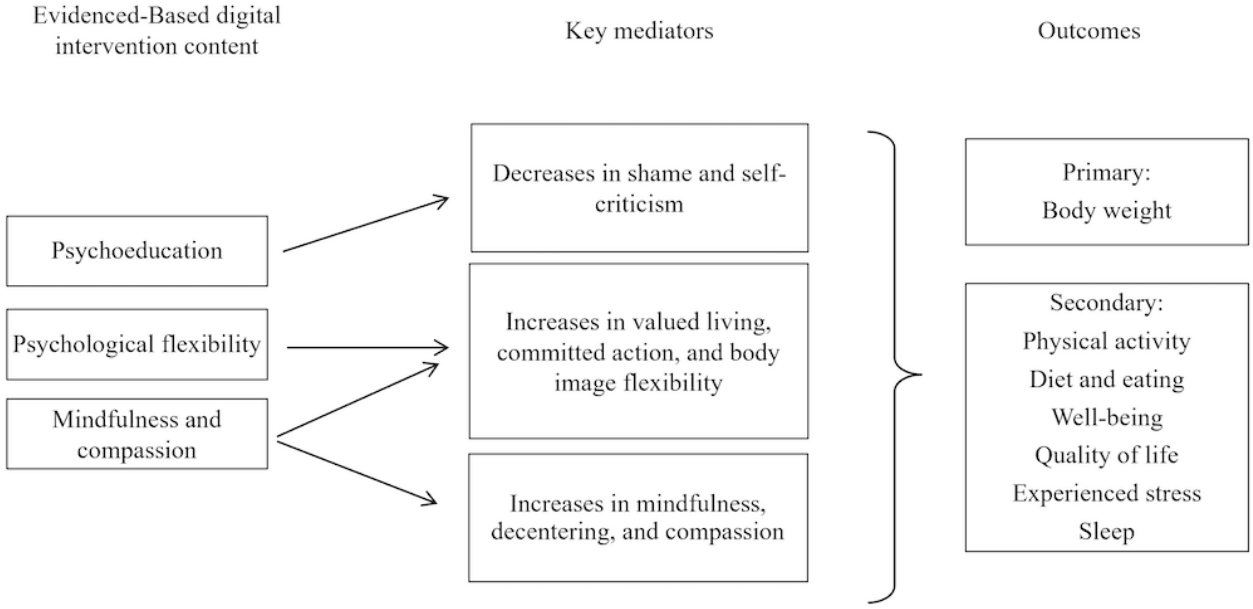
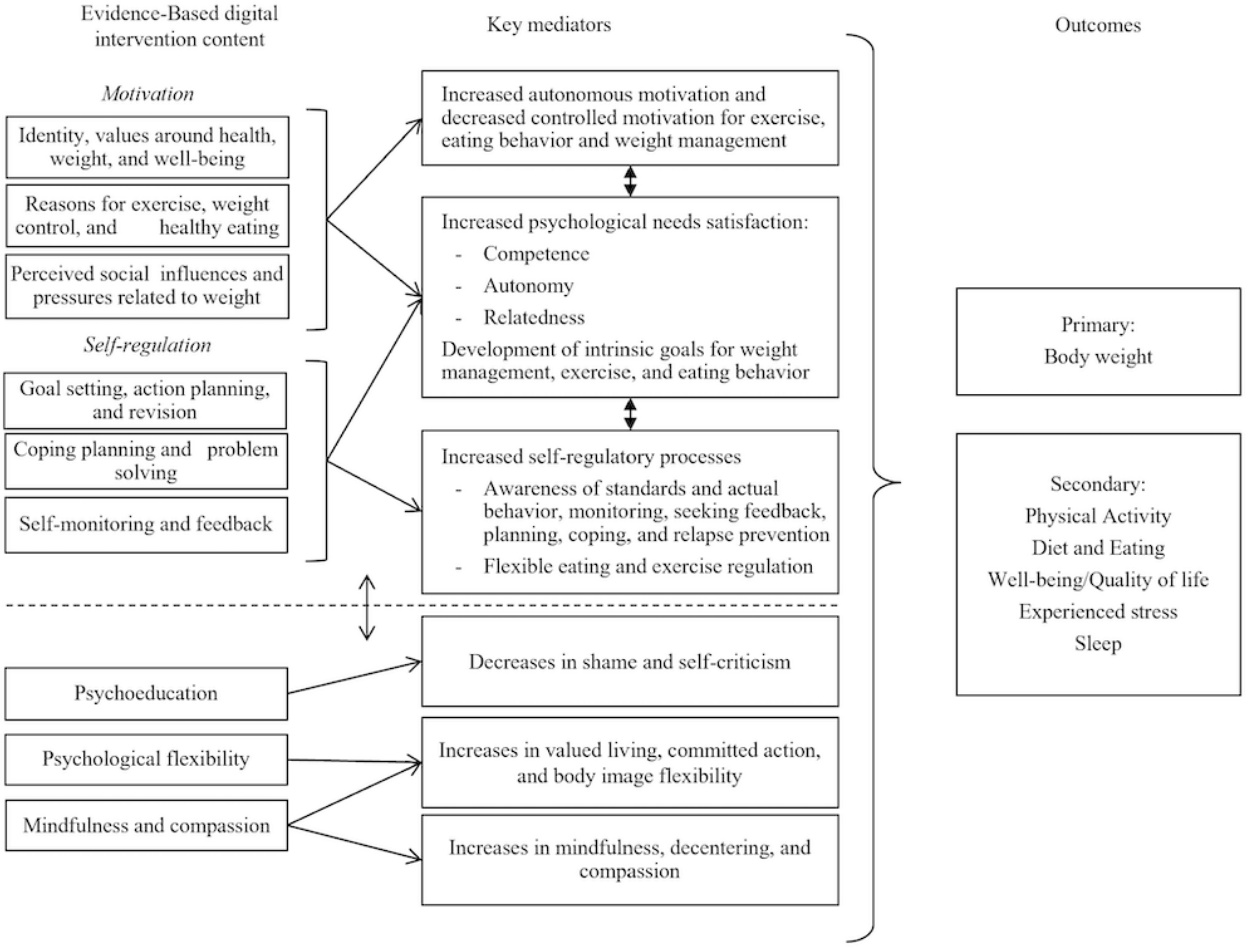


Figure 4. Logic model: combined arm.



Step 2: Selection of Behavior Change Techniques

For the *motivation and self-regulation* version of the toolkit, 29 intervention techniques were selected: 11 from the BCTT version 1, mainly focusing on goals, planning, and self-monitoring, and 18 from the classification of MBCTs, targeting autonomy, relatedness, and competence (see Multimedia Appendix 4 [26,27] for the full list of techniques). For the *emotion regulation* arm, 25 techniques were defined based on contextual behavioral science approaches, specifically,

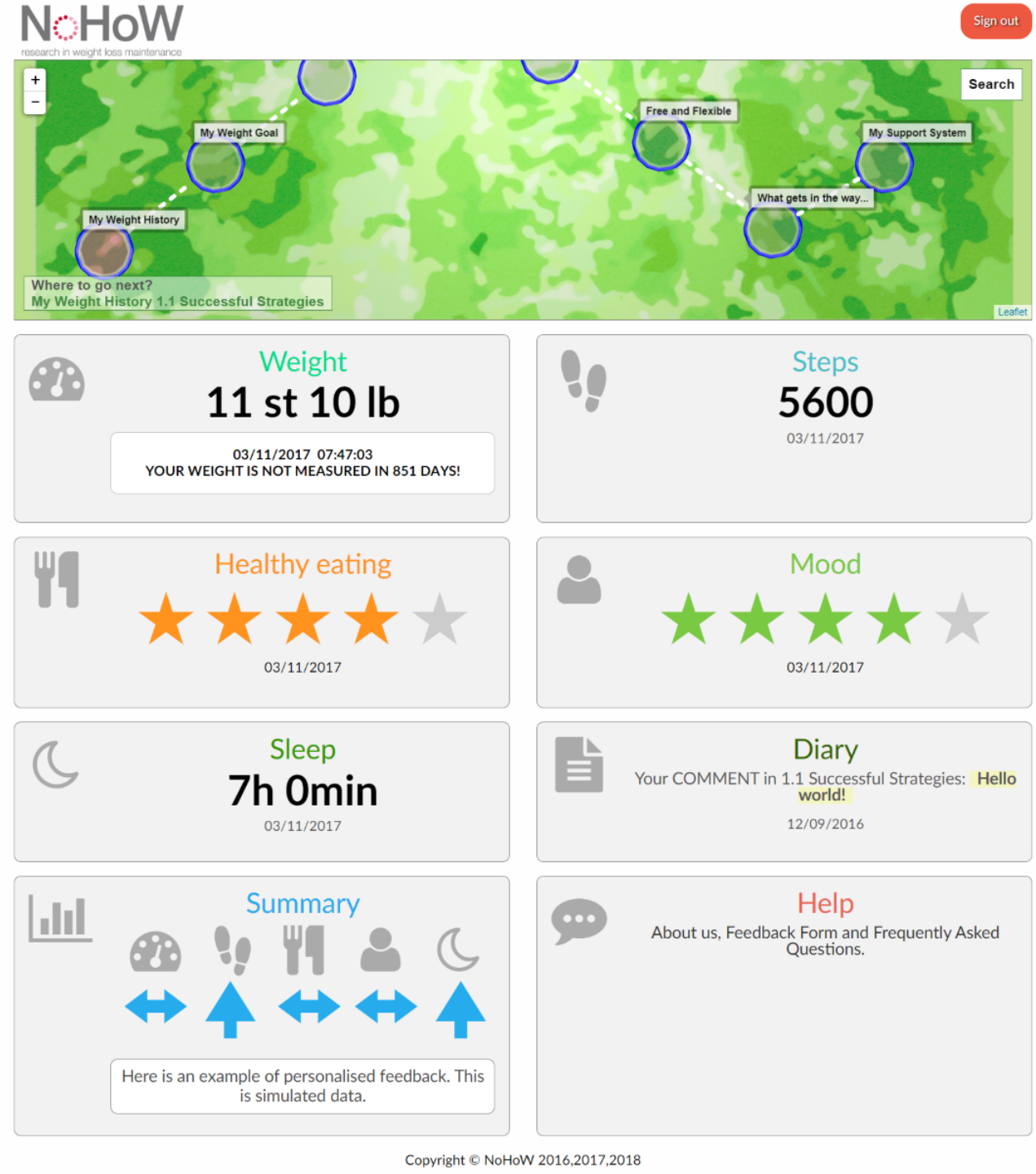
compassion-focused therapy, mindfulness-based interventions, and acceptance and commitment therapy (see Multimedia Appendix 5 for the full list of techniques).

Step 3: Translation of the Techniques Into the NoHoW Toolkit Components

Overview

The implementations of the techniques in the toolkit components are described below. When a user logged in to the toolkit, the main view (Figure 5) consisted of a tile-based dashboard, which was available for all toolkit users; a personal route map, which was only available for intervention users; weekly emails; a diary section (list of comments or reflections that a user could add at the end of each session); and a help section (information about the study, frequently asked questions, and instructions for navigating in the toolkit). The first 3 are described in detail next.

Figure 5. Navigating to a Healthy Weight Toolkit dashboard. st: stone; lb: pounds; h: hours; min: minutes



Tile-Based Dashboard

The dashboard tiles provided an at-a-glance view of the user's most recent data. The toolkit presented separate tiles on physical activity, sleep, and weight data retrieved from the Fitbit activity tracker (Charge 2) and the Fitbit smart Wi-Fi scale (Aria). All users could visualize their data for 1 week and for 1 month, 3 months, and 6 months. For the *motivation andself-regulation* arm and for the *combined* arm, the tiles also presented the user's goals and action plans as well as coping plans related to physical activity and diet. These plans could be reviewed at any time in the tiles. In addition, there were 2 tiles for mood and diet, each based, respectively, on the response to a single question, "How do you rate your mood, at the moment?" and "How satisfied are you with your diet, at the moment?" (5-star response scales). The tiles highlighted if no recent data existed, thus prompting continuation of monitoring. By clicking on a tile, the user could view long-term graphical feedback in different time scales and make self-assessments with a star rating feature (Multimedia Appendix 6).

Personal Map

The personal map (not available to the control arm) contained the modules of the toolkit, expressed in sessions (Multimedia Appendix 7). The sessions available depended on the arm to which the user was randomized. As the full content for each arm of the toolkit was available from the first day of use, it was considered useful to have a map to guide the user through the optimal theory-informed order of the content. When a user clicked on a module, the sessions of a module were displayed (Multimedia Appendix 7). There were regular sessions (lasting approximately 5 minutes) and short sessions (lasting 1 minute each), and the user was prompted to access a specific activity during the week. The *motivation and self-regulation* arm presented 17 regular sessions, organized in 8 modules. The estimated duration of these sessions ranged from 3 to 19 minutes. There were also 4 short sessions. The *emotion regulation* arm presented 17 regular sessions as well as 2 short ones distributed through 7 modules. The estimated duration ranged from 2 to 27 minutes. The *combined* arm presented all sessions from both arms distributed in 17 weekly modules.

Tables 2 and 3 describe the modules' sessions for all trial arms. The techniques selected were implemented in the sessions using various types of activities, or implementations, which are described in brief here (Multimedia Appendix 8):

1. Whiteboard animations: 1-minute videos that use animation to provide educational content. The toolkit presented videos about self-monitoring, self-confidence, myths and facts, flexibility and body image, autonomous motivation, evolutionary perspective of eating, conflicting messages, food functions, shame and self-criticism, obstacles to a healthier life, compassion, and mindfulness (Multimedia Appendix 9).
2. Questionnaires: interactive exercises to collect data from the user and provide immediate feedback (eg, a brief questionnaire concerning the reasons for personal ideal weight [motivation and self-regulation arm] or concerning compassionate attitudes [emotion regulation arm]; (Multimedia Appendix 10).
3. Testimonials: short personal stories describing, for example, the difficulties, challenges, and successes in a weight management process (Multimedia Appendix 11). Users could choose between a story of a male or female character.
4. Downloadable audios: audios containing mindfulness exercises, lasting approximately 5 minutes. They were only available in the emotion regulation and combined arms.

Table 2. Theme and goals of each module of the motivation and self-regulation arm.

Module (theme)	Core goals	Mechanisms of action
Module 1: My weight history	• Review weight change history: weight trajectory and characteristics of previous weight changes (strategies and feelings associated) • Identify weight loss strategies or approaches used in the last weight loss attempt and assess their sustainability in time	Self-awareness and competence
Module 2: My weight goal	• Learn about self-monitoring and self-referenced feedback and its role in goal-setting and revision, and reflect on individual options and preferences • Reflect on ideal and acceptable weights (and where these terms originated and what they mean to the person) • Understand the importance of setting self-relevant and optimal goals: set weight-related goal	Self-regulation capacity, competence, autonomy (ownership), and intrinsic (vs extrinsic) goals
Module 3: Myths and facts	• Promote factual knowledge about energy balance–related behaviors (exercise and diet) and WLM[a]	Competence and autonomous motivation
Module 4: My healthy goals	• Promote awareness of multiple choices around behavior changes and WLM in the long term (there is no “right” way) • Find individual interests and seek enjoyment and personal meaning around health behaviors • Explore personal resources for engaging in health behaviors (eg, skills)	Autonomy (perceived choice), autonomous motivation, competence, and self-regulation capacity
Module 5: My goals and values	• Prompt reflection on personal reasons for WLM (weight goals) and related behaviors by differentiating internal (autonomous) and external (controlled) motives and their relationship to sustained behavior change (sense of ownership) • Explore sources of body image or ideal (eg, societal norms, media, and significant others) and its consequences (motivation and well-being)	Intrinsic (vs extrinsic) goals (ie, life aspirations), autonomous versus controlled motivation, autonomy (ownership), and perceived social influences and pressures related to weight
Module 6: Free and flexible	• Identify functional and dysfunctional investment in body appearance (by exploring sources of body image or ideal (eg, societal norms, media, and significant others) and its consequences (motivation and well-being) and promote satisfaction with one’s body, at any size • Explore links between internal (feels free and choiceful) and external (feels pressured) motives and eating and exercise regulation (eg, rigid vs flexible approach)	Intrinsic (vs extrinsic) goals, autonomous versus controlled motivation, autonomy (ownership), and rigid versus flexible behavior regulation
Module 7: What gets in the way	• Identify challenges and barriers to current behavioral patterns and identify resources to increase capability to deal with them • Identify strategies to deal with these barriers (coping plans) and to focus attention on how behavior changes serve other important life goals	Competence, autonomous motivation, and self-regulation capacity
Module 8: My support system	• Identify sources of social support and reflect on what they mean (eg, pressured and conditional support vs unconditional support) • Increase skills in seeking social support and dealing with social and peer pressures (eg, assertiveness) • Explore reaching out to others as a role model or source of support and expertise in WLM	Relatedness, autonomous motivation, perceived social influences and pressures related to weight, and self-regulation processes

[a]WLM: weight loss maintenance.

Table 3. Theme and goals of each module of the emotion regulation arm.

Module (theme)	Core goals	Mechanisms of action
Module 1: Why do we eat?	• Promote the understanding of the difficulties in regulating eating behavior because our systems are not yet evolved to restrict eating behavior (it is not our fault) • Promote the knowledge about how our body works to stop people fighting against their bodies and begin working with them instead • Promote the understanding that food has multiple functions (it is not our fault) • Promote the reflection about the conflicting messages in our modern society about eating and physical activity and using food as a way of comfort and achieving a thin and fit body image • Clarify how these conflicting messages create additional stress and how eating may emerge as a way to soothe the self and manage stress	Evolutionary approaches to eating behavior and physical activity, deshaming, and emotion regulation
Module 2: Eating awareness	• Unveiling emotional and stress eating • Identify the traps in using food to regulate emotions and cope with stress • Learn how to eat mindfully • Promote awareness of satiety and hunger cues	Emotional eating, emotion regulation, and mindfulness
Module 3: Roadblocks to change	• Do we need shame and self-criticism to manage our weight? Understand the evolved functions of shame and self-criticism • Clarify the negative effects of self-criticism and shame on weight management, body image, and physical activity • Consolidate the inefficacy of shame and self-criticism to cope with stress and maintain changes • Promote creative hopelessness	Shame, self-criticism, creative hopelessness, and stress management
Module 4: Living a healthy life	• Foster creative hopelessness • Promote the clarification of values • Identify how life can be so disconnected from values • Identify the obstacles to a valued life • Discuss the role of avoidance and the control agenda as obstacles to a valued life • Create value-related goals and step-by-step actions • Encourage committed actions to values in daily life (ie, the importance of healthy eating patterns and physical activity)	Creative hopelessness, values, avoidance, control agenda, and committed action
Module 5: Learning to just be	• Introduce and reduce automatic pilot • Reflect on the needed shift from the doing mode to the being mode • Clarify what mindfulness is and evidence of its benefits • Increase awareness and acceptance of the present moment • Use the breath as an anchor for the present moment • Introduce the 3-minute breathing space as a way of being fully present with a different frame of mind • Learn how to use the 3-minute breathing space to deal with difficult emotions, sensations, thoughts, or the stress of daily life • Learn how to use the body as an anchor to the present moment experience • Increase awareness and acceptance of unwanted internal experiences (emotions and physical sensations) • Increase awareness of the body and the body in movement • Use mindfulness to take better care of your body • Increase awareness and acceptance of thoughts • Promote a defused perspective on thoughts: Thoughts are not facts • Promote an observer perspective of internal experiences • Promote the identification and awareness of stress responses in the body • Foster a more adaptive and healthy way to cope with stress and negative emotions	Mindfulness, automatic pilot, being mode, acceptance, awareness, coping, bodily awareness, mindful movement, decentering, and emotion regulation

Module (theme)	Core goals	Mechanisms of action
Module 6: Cultivating compassion	• Introduce loving-kindness • Understand how the need for compassion emerges from our evolved brain and emotional systems • Understand what compassion is • Clarify that compassion takes courage • Learn the basic skills of cultivating compassion • Learn how to prepare the body for the compassion practices • Develop and cultivate the qualities of the compassionate self • Understanding the role of interpersonal difficulties and how stigma generates stress • Cultivate compassion in my relationship with others • Identify the main reasons people report fearing compassion • Clarify what compassion is and what it is not • Discuss the paradoxical effects of fears of compassion • Help people to overcome their fears of compassion • Cultivate self-compassion (of one's body image, thoughts, and emotions) • Build the capacity for acting compassionately toward one's body • Take the courage to engage in hard but necessary actions: The importance of physical activity	Compassion, loving-kindness, postures, facial expressions and voice tones, soothing rhythm breathing, compassionate imagery, compassion for others, fears of compassion, and self-compassion
Module 7: Final destination: A new start	• Promote the early identification of relapse signals • Distinguish lapse from relapse • Draw an action plan to deal with the relapse • Prevent new relapses	Relapse prevention, mindfulness, and committed action

In addition, the toolkit presented 2 other features, as follows:

1. Extra support, which was triggered when individuals reached a threshold of weight regain >3% above their target weight. This extra module stayed available until users returned to a weight interval of ±3% (ie, the interval defined as weight maintenance). For users in the *motivation and self-regulation* arm and the *emotion regulation* arm, this support further directed the user to content useful for coping with relapses (Multimedia Appendix 12).

2. Individualized feedback, which was included as a feature in the *motivation and self-regulation* arm and in which time series data retrieved over 1-2 months for each participant (activity and sleep patterns from Fitbit devices, daily and weekly body weight measures taken by the Fitbit Aria Wi-Fi scales, and use of the toolkit) were analyzed to examine weight change patterns. There were also indicators of weekly versus weekend and earlier versus later in the week used as possible predictors to alert participants to weekly cycles in weight change. Feedback could only be provided if sufficient data were collected (at least 30 days and variation in weight) and consisted of short messages such as "For you, being active or exercising seems linked to better weight management" or "Your weight management seems better on weekdays" that were displayed on the homepage.

Weekly Emails

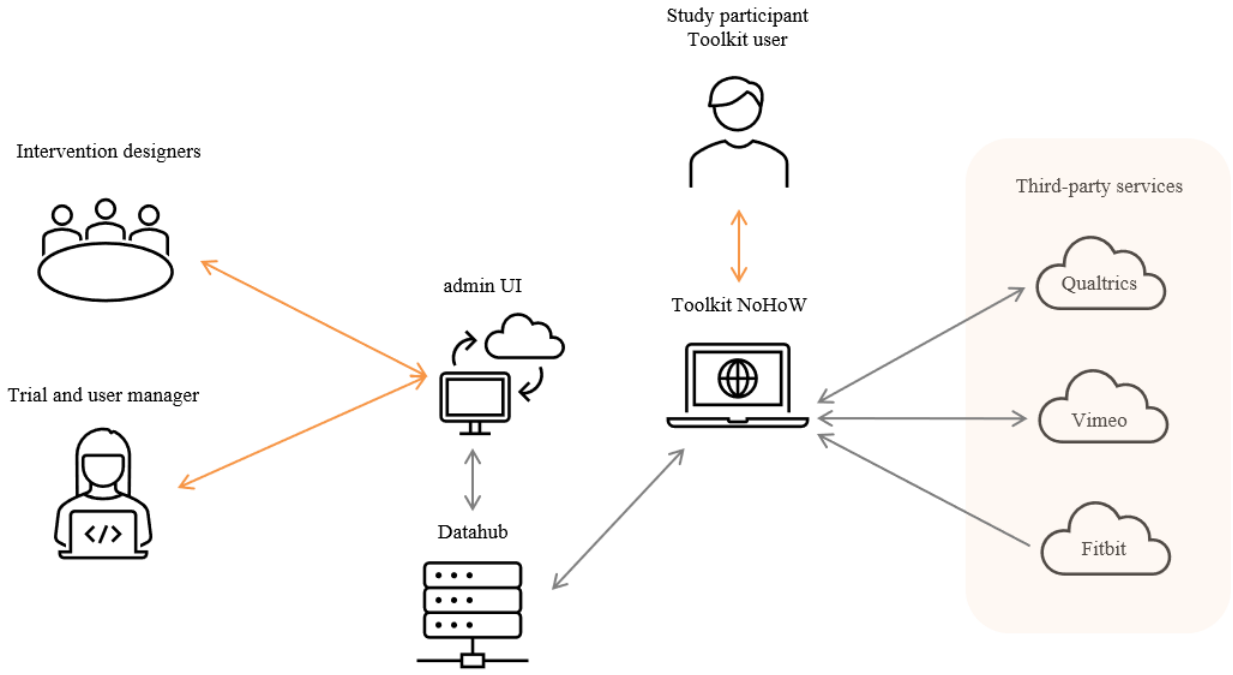
During the intervention period, weekly emails were sent to the control arm, prompting the users to access the dashboard. These emails also included links to general information about weight management (eg, guidelines for healthy lifestyles from government or scientific association websites).

The other arms also received weekly emails prompting the users to access a specific session in the toolkit suggested for that week, that is, the participants from the *motivation and self-regulation* arm received the link for a session developed with that rationale, whereas the participants from the *emotion regulation* arm received the link to an emotion regulation–specific session.

Step 4: Toolkit Technical Development

Figure 6 depicts the final overall architecture of the NoHoW Toolkit and its connections to other systems required to deliver the intervention. The toolkit consisted of 2 layers. The front end consisted of an HTML skeleton with JavaScript modules. A responsive web design approach and mobile-first concept was used in the development. The back end responds to front end requests, communicates with the MySQL database on the datahub server, and accesses Fitbit data on Fitbit's server. The datahub stores and analyzes data, providing personalized feedback to users by relaying information back to users through the personalized feedback panels on the dashboard of the toolkit. The intervention content was delivered through text, images, videos, questionnaires, and interactive exercises. The text and images were implemented as HTML. The videos were hosted on Vimeo (Vimeo, Inc) and embedded in the intervention sessions. The embedded questionnaires were implemented with Qualtrics (Qualtrics XM), which was also used for managing the study questionnaires in the RCT. New interfaces were built to retrieve questionnaire answers from Qualtrics and visualize them in the toolkit front end to enable exercises where instant feedback was needed. Interactive exercises were tailor made as mini apps. One example of a mini app was an exercise enabling the users to visualize key moments in their personal weight management history as a graph.

Figure 6. The overall architecture of the Navigating to a Healthy Weight Toolkit and its connections to third-party systems. NoHoW: Navigating to a Healthy Weight; UI: user interface.



Each user was provided with a user account for the toolkit website as well as a Fitbit account to gather activity, sleep, and weight data. The personal map was designed iteratively with the help of a UX designer. First, 3 different concepts for visualizing the personal map were designed, and paper prototypes illustrating the concepts in different stages of use were created. The prototypes were evaluated with 6 participants in 2 evaluation rounds, first with 3 participants, after which major usability issues were fixed. Next, the updated prototypes were evaluated with another 3 participants. The evaluation consisted of 45-minute face-to-face sessions with each of the participants, who were asked to complete different tasks with the prototypes and provide feedback on the usability and visual appearance of the concepts. Finally, the participants were asked to rate the concepts as the best, second best, and worst. A concept illustrating a path on a map received the highest ratings and was selected. The final design was created by incorporating user feedback and some highly rated features from the other 2 concepts. All intervention modules were displayed on the main level of the map, and completed modules were highlighted in pink; modules that had not been completed were highlighted in a lighter shade of pink. When a user clicked on a module, the sessions of that module were displayed. This level of the map also indicated which of the sessions had been completed. By clicking on a session, the user gained access to the intervention content. On both levels of the map, there was also a box at the bottom left of the screen, indicating the next recommended session and providing a shortcut to the session.

The admin UI served intervention developers and trial managers. The most important component in terms of the intervention was the content management system (CMS) because of the large volume of intervention content and the need to constantly refine and update it during the iterative development of the toolkit. Thus, enabling the intervention designers to manage the content facilitated the collaboration between the software developers and the intervention designers. The CMS enabled the intervention designers to describe the intervention structure, content items, and their relationships to each other with a web-based UI without programming expertise. The CMS is based on the Django CMS (Django CMS Association), but its functionality has been tailored to meet the needs of the intervention design.

Step 5: User Evaluation and Toolkit Refinement

A total of 37 eligible participants expressed interest in participating in the first study. Of the 37 participants, 20 (54%) started using the toolkit, 14 (38%) completed the 5-day questionnaire, and 9 (24%) completed the 30-day questionnaire. The average age of the participants was 41.4 (SD 11.7; range 23-63) years, and of the 37 participants, 22 (59%) were women. Although retention was low, the collected data were sufficient to allow for the analysis of the feasibility study. The results from the study were used to revise the toolkit specifications. The main threats to engagement and acceptance of the toolkit were identified, and several improvements were made regarding UX design, content development, technical performance, and usability (Multimedia Appendix 2). For example, to facilitate navigation in the toolkit, the personal map component was included, videos were refilmed to make them more energetic and less clinical, and technical performance and usability were improved by ensuring scaling of the toolkit to different screen sizes and ensuring compatibility with different browsers and mobile platforms. Ethical approval was obtained from the Psychology Research Ethics Committee of the University of Derby.

The qualitative study assessing users' needs support when engaging with toolkit version 2.0 had a sample size of 12 adults aged 23-57 years (women: 9/12, 75%) who lost between 10% and 43% of their weight before taking part in the qualitative study. Most of the participants had experience in using digital tools, mainly to monitor their diet and physical activity. When testing the toolkit, the participants did not report difficulties, and the fact that the toolkit could be used on both a computer and mobile phone was considered an advantage. The participants considered that the self-monitoring tools were positive elements and considered that engaging in the toolkit sessions for 5-10 minutes per week was a reasonable time commitment. From the interviews, it was concluded that all participants felt that the toolkit content was acceptable; in particular, the whiteboard videos were considered to be very appealing and engaging in terms of both the design and content, with a participant stating, "I think I prefer the drawing hand, in my opinion the hand is more engaging, it has a lot of movement." There were no major issues or suggestions for changes from the interviews (Multimedia Appendix 3). Approval from the research ethics committees at the University of Lisbon was obtained before recruitment and data collection.

The final version of the NoHoW Toolkit was presented in 4 versions available for each of the trial arms as well as in 3 languages for each of the trial centers. All versions included a tile-based dashboard displaying data from digital tracking devices (Fitbit Charge 2 wrist device and Aria smart scale) for weight, steps, and sleep, as well as manually entered data for eating and mood. It also included a diary section and a help section. In addition, the 3 versions available to each of the intervention arms—*motivation and self-regulation*, *emotion regulation*, and *combined*—displayed a personal map containing the modules of the toolkit. The final toolkit comprised 34 sessions, distributed through 15 modules, with active content being presented for 4 months. The *motivation and self-regulation* arm consisted of 17 sessions, distributed through 8 modules; the *emotion regulation* arm was presented with 17 sessions distributed through 7 modules; and the *combined* arm received the full toolkit. A detailed logging of user interactions with the toolkit was implemented to enable determining how often the users visited the toolkit and how much time they spent in the toolkit and in the intervention sessions.

Discussion

Principal Findings

The NoHoW Toolkit was one of the first theory and evidence-based digital approaches for WLM that was systematically developed using standardized guidance, an interdisciplinary approach, design principles, and user testing. In all, 4 toolkit versions were designed and formally evaluated through a 2×2 factorial RCT. In the RCT, the control version of the toolkit only presented self-monitoring information on weight and physical activities, whereas the 3 intervention arms of the toolkit presented additional modular content targeting (1) motivational factors of behavior change, including the promotion of autonomous motivation (vs controlled motivation), intrinsic goals, and flexible regulation to change behaviors and maintain weight loss and self-regulation skills; the setting and review of optimal goals; action and coping planning; and action control; (2) emotion regulation and stress management to reduce weight-related shame and self-criticism, reduce difficulties in emotion regulation, and increase psychological flexibility, mindfulness, and compassion; or (3) both 1 and 2. These are factors that have been shown to have an impact on energy balance–related behaviors (physical activity and healthy diet), weight changes, and well-being [12-15,35-38].

The systematic approach used in the development of the NoHoW Toolkit—making use of guidance for complex behavior change interventions and taxonomies—was a core contributor to the best reporting, replication, and accumulation of evidence on effective behavioral approaches to WLM. The logic models created for the NoHoW Toolkit allow for statistical modeling of the components of the intervention, contributing to understanding what works in the context of long-term weight management. Furthermore, using explicit theory when developing interventions allows for the identification of the factors that influence the target behavior, the mechanisms of actions that operate along the pathway to change, and the best techniques to influence these factors, as well as an understanding of how these can affect engagement. In addition, the use of digital tools and engagement metrics (eg, real-time data on weight or activity changes and use patterns of the intervention content) can contribute to improving our understanding of the behavioral dynamic patterns as well as theory testing and refinement. Users were engaged in the design and iteration of the toolkit at several stages, and the findings from these studies informed several improvements in the design of the toolkit and its content.

Challenges and Lessons Learned

A challenge we faced in the development of the NoHoW Toolkit concerned the integration of the motivation and self-regulation theoretical approaches with emotion regulation to be tested in the 2×2 factorial trial. As there was no theory regarding, or empirical support for, the interactions between these approaches, the *combined* arm provided access to all the toolkit content available, and the users were prompted each week to access the emotion regulation content and the motivation and self-regulation content. In addition, there was some overlap among the sessions from the different approaches, for example, the sessions on values and on flexible eating. Furthermore, the content of the activities provided in the toolkit mainly focused on maintenance of weight loss by, for example, addressing the myths and facts of WLM, asking users to indicate the strategies they used to lose weight and how these strategies could fit with a long-term approach, and offering activities that addressed autonomous motivation as well as content that focused on mindfulness and compassion as ways to prevent relapses.

Although we used up-to-date frameworks to guide our choices for the content, structure, and flow of the toolkit, the final decisions on which components and features were to be included and how these would be implemented drew on the research team's judgment through iterative discussions that were informed by user testing and feedback. It should be noted that the decisions made in the development process were not

systemically justified and documented. Current frameworks and guidance provide little information on how to translate behavior change techniques into digital content (features). For example, what is the best way to implement a modeling technique in a digital intervention for it to be most effective? Should video or text or audio be used? Or should a combination of these be used? Behavior change techniques do not influence only the intervention content, but also the functionality of that content: how users interact with, and use, the app. This in turn affects how behavior change techniques can be translated into individual functionalities and how these functionalities form an overall concept that creates value for the user. Efforts are currently ongoing to develop tools to support researchers and interventionists in standardizing reporting to inform intervention design, content, and delivery [39,40].

The development of the NoHoW Toolkit involved close collaboration between behavioral scientists and web developers. The toolkit (and hence intervention) design was developed at the beginning of the project and limited by time and resource constraints. Where possible, development was accelerated by using existing commercial solutions, such as implementing intervention exercises using Qualtrics questionnaires and using commercially available tracking devices and their data aggregation capabilities. The resource constraints meant negotiating acceptable compromises between the teams in the implementation of some toolkit components. It was difficult to evaluate beforehand how much work each functionality required and how essential it actually was, considering the evidence base.

It can be challenging to develop a totally new digital behavior change intervention in a time-constrained project that includes a large-scale effectiveness evaluation. Although users were involved in the iterative development of the toolkit, it was not possible to simulate long-term, real-life experiences of users of such a large entity in qualitative interviews or even in a 1-month feasibility study. This compromise is also relevant when judging if the toolkit was suitable for participants from different socioeconomic status and literacy levels. Future studies should consider adopting methodologies such as coparticipatory design or formative research to improve the suitability of the app content for diverse populations. Another limitation related to this was that user testing was not conducted with Danish users.

The UX data collected during the NoHoW RCT will be valuable in understanding how users' experiences change during the use of a long-term digital intervention. In future studies, when a complex long-term intervention is involved, it might be worthwhile to conduct a longer feasibility study to enable observation of the changes in user engagement and experience and use qualitative methods to investigate the reasons. Furthermore, an adaptive trial design might be a good alternative for testing and comparing different designs.

The schedule for the development was tight, which led to pressure to use existing commercial platforms in the development. Although including commercial options such as Fitbit can be an advantage because we can test real-world solutions, it also means less control over the features that are regularly added to the Fitbit app and trackers, which can contradict, or be in conflict or overlap with, the content provided in the NoHoW Toolkit, for example, the built-in *breathe* exercise in the Fitbit tracker was not aligned with the relaxation techniques recommended in the toolkit. Future interventions may favorably control these features by using devices used for research or codeveloping these with commercial partners, for example, by integrating new features on the Fitbit native app that would be released to the NoHoW participants only. Furthermore, developing a web app instead of a native mobile app limited the functionality and interaction mechanisms from the intervention design point of view. Focusing on the web-based app saved the trouble of developing separate apps for multiple platforms (Android and iOS), and the related CMS enabled updating of the intervention content in real time.

Although a factorial RCT allows a rigorous testing and comparison of different theory-based components, it also presents constraints regarding the continuous optimization of content to the individual that may arise during the trial (ie, the content could not be updated during the full length of the NoHoW trial). The use of optimization designs for digital interventions (eg, sequential multiple assignment randomized trial [41] or microrandomized trial [42] approaches) can support the identification and delivery of the content of interventions through the modeling of causal and time-varying effects, providing a more personalized and dynamic approach to complex behavior change in the context of long-term weight management.

Considering these challenges, it is likely that the interaction between academia and industry will be crucial to the development of digital behavior change solutions in the years to come. The industry has technical capabilities and capacity that could potentially synergize with academic efforts in developing rigorous, systematic, and dynamic behavior change interventions grounded in sound theoretical principles.

Conclusions

The NoHoW Toolkit is a theory-based web app aimed at supporting individuals who have already lost weight and are trying to manage their weight in the long term. It was systematically developed using standardized guidance, an interdisciplinary approach, design principles, and user testing, and it was formally evaluated through a 2×2 factorial RCT. It presents modular content on motivational factors of behavior change as well as behavior regulation and emotion regulation skills, and it integrates data from activity trackers and weight scales. The development of the toolkit involved a multidisciplinary and international team of experts who used a systematic and rigorous approach to derive its technical specifications and content limitations. The toolkit has been tested in the context of a multicountry randomized trial; in addition to the main trial results, the results from the mediation-moderation analysis, guided by the logic models, will provide further information about the most effective and engaging components for WLM, thereby contributing to the optimization of the NoHoW Toolkit in future deployments.

Acknowledgments

The authors thank Heather Patrick, Kirby Sainsbury, Paul Sacher, and Sarah Scott for their valuable inputs to the toolkit development and evaluation.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 643309. The materials presented and views expressed here are the responsibility of the authors only. The European Commission takes no responsibility for any use made of the information in this study.

Authors' Contributions

RJS, BLH, PJT, FFS, ME, and GH conceived the NoHoW project. MMM, ALP, PJT, EM, PV, MM, CD, GH, ME, and JL designed the toolkit. All authors contributed to the development of the content and functionalities of the toolkit. MH, EM, and RJS designed the user evaluation in the United Kingdom. MMM, ALP, and JE conducted the Portuguese usability study. MMM, ALP, JE, EM, MH, CD, and MM drafted the manuscript. MMM and EM contributed equally to this work. All authors reviewed and approved the manuscript.

Conflicts of Interest

RJS consults for Slimming World through Consulting Leeds, which is a wholly owned subsidiary of the University of Leeds. MMM and GH have previously consulted for Slimming World. Slimming World is a former partner in the Navigating to a Healthy Weight project. All other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Summary of a systematic review submitted as part of a deliverable in the Navigating to a Healthy Weight Project.
[PDF File (Adobe PDF File), 448 KB - jmir_v23i12e25305_app1.pdf]

Multimedia Appendix 2
Summary of a feasibility study of the NoHoW Toolkit submitted as part of a deliverable in the NoHoW Project. NoHoW: Navigating to a Healthy Weight.
[PDF File (Adobe PDF File), 83 KB - jmir_v23i12e25305_app2.pdf]

Multimedia Appendix 3
Summary of a qualitative analysis of the support needs and usability of the NoHoW Toolkit submitted as part of a deliverable in the NoHoW Project. NoHoW: Navigating to a Healthy Weight.
[PDF File (Adobe PDF File), 147 KB - jmir_v23i12e25305_app3.pdf]

Multimedia Appendix 4
Theoretical guiding principles underlying the intervention design: motivation and behavior regulation arm.
[PDF File (Adobe PDF File), 86 KB - jmir_v23i12e25305_app4.pdf]

Multimedia Appendix 5
Theoretical guiding principles underlying the intervention design: emotion regulation.
[PDF File (Adobe PDF File), 83 KB - jmir_v23i12e25305_app5.pdf]

Multimedia Appendix 6
Graphical view of a tile: weight tile.
[PDF File (Adobe PDF File), 152 KB - jmir_v23i12e25305_app6.pdf]

Multimedia Appendix 7
Personal route map.
[PDF File (Adobe PDF File), 105 KB - jmir_v23i12e25305_app7.pdf]

Multimedia Appendix 8
Motivation and self-regulation modules.
[PDF File (Adobe PDF File), 227 KB - jmir_v23i12e25305_app8.pdf]

Multimedia Appendix 9
Example of a whiteboard animation.

[PDF File (Adobe PDF File), 195 KB - jmir_v23i12e25305_app9.pdf]

Multimedia Appendix 10
Example of a questionnaire.
[PDF File (Adobe PDF File), 90 KB - jmir_v23i12e25305_app10.pdf]

Multimedia Appendix 11
Example of a testimony.
[PDF File (Adobe PDF File), 129 KB - jmir_v23i12e25305_app11.pdf]

Multimedia Appendix 12
Extra support feature.
[PDF File (Adobe PDF File), 206 KB - jmir_v23i12e25305_app12.pdf]

References

1. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
2. Wadden TA, Tronieri JS, Butryn ML. Lifestyle modification approaches for the treatment of obesity in adults. Am Psychol 2020 Feb;75(2):235-251 [FREE Full text] [doi: 10.1037/amp0000517] [Medline: 32052997]
3. Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, et al. Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement. Health Technol Assess 2004 May;8(21):iii-iv, 1 [FREE Full text] [doi: 10.3310/hta8210] [Medline: 15147610]
4. Paixão C, Dias CM, Jorge R, Carraça EV, Yannakoulia M, de Zwaan M, et al. Successful weight loss maintenance: a systematic review of weight control registries. Obes Rev 2020 May;21(5):e13003. [doi: 10.1111/obr.13003] [Medline: 32048787]
5. Montesi L, El Ghoch M, Brodosi L, Calugi S, Marchesini G, Dalle GR. Long-term weight loss maintenance for obesity: a multidisciplinary approach. Diabetes Metab Syndr Obes 2016 Feb;9:37-46 [FREE Full text] [doi: 10.2147/DMSO.S89836] [Medline: 27013897]
6. Stubbs R, Lavin J. The challenges of implementing behaviour changes that lead to sustained weight management. Nutr Bull 2013 Feb 07;38(1):5-22 [FREE Full text] [doi: 10.1111/nbu.12002]
7. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
8. Park S, Hwang J, Choi Y. Effect of mobile health on obese adults: a systematic review and meta-analysis. Healthc Inform Res 2019 Jan;25(1):12-26 [FREE Full text] [doi: 10.4258/hir.2019.25.1.12] [Medline: 30788177]
9. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
10. 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(sep29 1):a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
11. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 07;348(mar07 3):g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
12. Dombrowski S, Sniehotta F, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2012 Mar;6(1):7-32 [FREE Full text] [doi: 10.1080/17437199.2010.513298]
13. Sniehotta FF, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol 2005 Jul;35(4):565-576. [doi: 10.1002/ejsp.258] [Medline: 25855820]
14. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
15. Teixeira PJ, Mata J, Williams GC, Gorin AA, Lemieux S. Self-regulation, motivation, and psychosocial factors in weight management. J Obes 2012;2012:582348 [FREE Full text] [doi: 10.1155/2012/582348] [Medline: 23209888]
16. Ntoumanis N, Ng JY, Prestwich A, Quested E, Hancox JE, Thøgersen-Ntoumani C, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior,

physical, and psychological health. Health Psychol Rev 2021 Jun 03;15(2):214-244. [doi: 10.1080/17437199.2020.1718529] [Medline: 31983293]
17. Sniehotta FF, Evans EH, Sainsbury K, Adamson A, Batterham A, Becker F, et al. Behavioural intervention for weight loss maintenance versus standard weight advice in adults with obesity: a randomised controlled trial in the UK (NULevel Trial). PLoS Med 2019 May 7;16(5):e1002793 [FREE Full text] [doi: 10.1371/journal.pmed.1002793] [Medline: 31063507]
18. Duarte C, Stubbs J, Pinto-Gouveia J, Matos M, Gale C, Morris L, et al. The impact of self-criticism and self-reassurance on weight-related affect and well-being in participants of a commercial weight management programme. Obes Facts 2017 Apr 4;10(2):65-75 [FREE Full text] [doi: 10.1159/000454834] [Medline: 28376484]
19. Bowen S, Chawla N, Witkiewitz K. Mindfulness-based relapse prevention for addictive behaviors. In: Baer RA, editor. Mindfulness-Based Treatment Approaches.2nd ed. Cambridge, MA: Academic Press; 2014:141-157.
20. Hendershot CS, Witkiewitz K, George WH, Marlatt GA. Relapse prevention for addictive behaviors. Subst Abuse Treat Prev Policy 2011 Jul 19;6:17 [FREE Full text] [doi: 10.1186/1747-597X-6-17] [Medline: 21771314]
21. O'Reilly GA, Cook L, Spruijt-Metz D, Black DS. Mindfulness-based interventions for obesity-related eating behaviours: a literature review. Obes Rev 2014 Jun 18;15(6):453-461 [FREE Full text] [doi: 10.1111/obr.12156] [Medline: 24636206]
22. Scott SE, Duarte C, Encantado J, Evans EH, Harjumaa M, Heitmann BL, et al. The NoHoW protocol: a multicentre 2×2 factorial randomised controlled trial investigating an evidence-based digital toolkit for weight loss maintenance in European adults. BMJ Open 2019 Sep 30;9(9):e029425 [FREE Full text] [doi: 10.1136/bmjopen-2019-029425] [Medline: 31575569]
23. 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]
24. Bartholomew L, Markham C, Ruiter R, Fernández M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach. Hoboken, New Jersey, United States: John Wiley & Sons; 2016.
25. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
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;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
27. Teixeira PJ, Marques MM, Silva MN, Brunet J, Duda JL, Haerens L, et al. A classification of motivation and behavior change techniques used in self-determination theory-based interventions in health contexts. Motiv Sci 2020 Dec 05;6(4):438-455. [doi: 10.1037/mot0000172]
28. Santos I, Sniehotta FF, Marques MM, Carraça EV, Teixeira PJ. Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. Obes Rev 2017 Jan 21;18(1):32-50 [FREE Full text] [doi: 10.1111/obr.12466] [Medline: 27653242]
29. Teixeira PJ, Carraça EV, Marques MM, Rutter H, Oppert J, De Bourdeaudhuij I, et al. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. BMC Med 2015 Apr 16;13:84 [FREE Full text] [doi: 10.1186/s12916-015-0323-6] [Medline: 25907778]
30. Vartanian LR, Porter AM. Weight stigma and eating behavior: a review of the literature. Appetite 2016 Jul 01;102:3-14. [doi: 10.1016/j.appet.2016.01.034] [Medline: 26829371]
31. Tomiyama AJ. Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model. Appetite 2014 Nov;82:8-15. [doi: 10.1016/j.appet.2014.06.108] [Medline: 24997407]
32. Gibson E. The psychobiology of comfort eating: implications for neuropharmacological interventions. Behav Pharmacol 2012 Sep;23(5-6):442-460. [doi: 10.1097/FBP.0b013e328357bd4e] [Medline: 22854304]
33. Puhl RM, Brownell KD. Confronting and coping with weight stigma: an investigation of overweight and obese adults. Obesity (Silver Spring) 2006 Oct;14(10):1802-1815 [FREE Full text] [doi: 10.1038/oby.2006.208] [Medline: 17062811]
34. Dallman MF, Pecoraro N, Akana SF, La Fleur SE, Gomez F, Houshyar H, et al. Chronic stress and obesity: a new view of "comfort food". Proc Natl Acad Sci U S A 2003 Sep 30;100(20):11696-11701 [FREE Full text] [doi: 10.1073/pnas.1934666100] [Medline: 12975524]
35. Carrière K, Khoury B, Günak MM, Knäuper B. Mindfulness-based interventions for weight loss: a systematic review and meta-analysis. Obes Rev 2018 Feb;19(2):164-177 [FREE Full text] [doi: 10.1111/obr.12623] [Medline: 29076610]
36. Ruffault A, Czernichow S, Hagger MS, Ferrand M, Erichot N, Carette C, et al. The effects of mindfulness training on weight-loss and health-related behaviours in adults with overweight and obesity: a systematic review and meta-analysis. Obes Res Clin Pract 2017 Sep;11(5 Suppl 1):90-111. [doi: 10.1016/j.orcp.2016.09.002] [Medline: 27658995]
37. Forman EM, Butryn ML. A new look at the science of weight control: how acceptance and commitment strategies can address the challenge of self-regulation. Appetite 2015 Jan;84:171-180 [FREE Full text] [doi: 10.1016/j.appet.2014.10.004] [Medline: 25445199]
38. Lillis J, Wing R. Behavioral strategies in weight management. In: Treatment of the Obese Patient. Switzerland: Springer; 2014.

39. Michie S, West R, Finnerty AN, Norris E, Wright AJ, Marques MM, et al. Representation of behaviour change interventions and their evaluation: development of the upper level of the behaviour change intervention ontology. Wellcome Open Res 2020 Jun 10;5:123. [doi: 10.12688/wellcomeopenres.15902.1]
40. Marques MM, Carey RN, Norris E, Evans F, Finnerty AN, Hastings J, et al. Delivering behaviour change interventions: development of a mode of delivery ontology. Wellcome Open Res 2020 Jun 10;5:125. [doi: 10.12688/wellcomeopenres.15906.1]
41. 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 Aug 05;11(4):426-434 [FREE Full text] [doi: 10.1177/1740774514536795] [Medline: 24902922]
42. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S(Suppl):1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]

Abbreviations

BCTT: Behavior Change Techniques Taxonomy
CMS: content management system
MBCT: Motivation and Behavior Change Techniques
NoHoW: Navigating to a Healthy Weight
RCT: randomized controlled trial
UI: user interface
UX: user experience
WLM: weight loss maintenance

Edited by R Kukafka; submitted 27.10.20; peer-reviewed by H Mehdizadeh, I Shubina, O Borek, N Buckland, C Zhang; comments to author 18.12.20; revised version received 08.02.21; accepted 16.03.21; published 03.12.21.

Please cite as:
Marques MM, Matos M, Mattila E, Encantado J, Duarte C, Teixeira PJ, Stubbs RJ, Sniehotta FF, Ermes M, Harjumaa M, Leppänen J, Välkkynen P, Silva MN, Ferreira C, Carvalho S, Palmeira L, Horgan G, Heitmann BL, Evans EH, Palmeira AL
A Theory- and Evidence-Based Digital Intervention Tool for Weight Loss Maintenance (NoHoW Toolkit): Systematic Development and Refinement Study
J Med Internet Res 2021;23(12):e25305
URL: https://www.jmir.org/2021/12/e25305
doi:10.2196/25305
PMID:34870602

Original Paper

A Theory-Based, Multidisciplinary Approach to Cocreate a Patient-Centric Digital Solution to Enhance Perioperative Health Outcomes Among Colorectal Cancer Patients and Their Family Caregivers: Development and Evaluation Study

Su Wei Wan[1,2], BSc; Choon Seng Chong[2,3,4], MBBS, MRCS, MMed, FRCS; Ee-Lin Toh[4,5,6], MBBS, MRCSEd, MMed, FRCSEd; Siew Hoon Lim[7], BSc, PhD; Carol TT Loi[8], MSc, PhD, Grad Dip; Yuen Foong Henry Lew[9], BA, MA; Matthew Chin Heng Chua[10], PhD; Xin Pei Jee[2,11], Diploma in Nursing; Guangyu Liu[1], BSc; Lixia Zhu[1], BSc, MSc, PhD; Minna Pikkarainen[12,13,14], PhD; Hong-Gu He[1,2], BSc, MSc, PhD

[1]Alice Lee Centre for Nursing Studies, National University of Singapore, Singapore, Singapore

[2]National University Health System, Singapore, Singapore

[3]Division of Colorectal Surgery and Surgical Oncology, Department of Surgery, National University Cancer Institute, National University Hospital, Singapore, Singapore

[4]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[5]Department of Colorectal Surgery, Singapore General Hospital, Singapore, Singapore

[6]Duke-NUS Medical School, Singapore, Singapore

[7]Division of Nursing, Singapore General Hospital, Singapore, Singapore

[8]Centre of Psychosocial Support, Singapore Red Cross Academy, Singapore, Singapore

[9]Department of Psychology, School of Humanities and Behavioural Science, Singapore University of Social Sciences, Singapore, Singapore

[10]Medical and Cybernetics Systems, Institute of Systems Science, National University of Singapore, Singapore, Singapore

[11]Division of Colorectal Surgery, Department of Nursing, National University Hospital, Singapore, Singapore

[12]Faculty of Health Sciences and Faculty of Technology, Art and Design, Oslomet - Oslo Metropolitan University, Oslo, Norway

[13]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden

[14]Faculty of Medicine and Oulu Business School, University of Oulu, Oulu, Finland

Corresponding Author:
Hong-Gu He, BSc, MSc, PhD
Alice Lee Centre for Nursing Studies
National University of Singapore
Level 2, Clinical Research Centre
Block MD11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 65167448
Fax: 65 67767135
Email: nurhhg@nus.edu.sg

Abstract

Background: Elective colorectal cancer (CRC) surgeries offer enhanced surgical outcomes but demand high self-efficacy in prehabilitation and competency in self-care and disease management postsurgery. Conventional strategies to meet perioperative needs have not been pragmatic, and there remains a pressing need for novel technologies that could improve health outcomes.

Objective: The aim of this paper was to describe the development of a smartphone-based interactive CRC self-management enhancement psychosocial program (iCanManage) in order to improve health outcomes among patients who undergo elective CRC surgeries and their family caregivers.

Methods: A multidisciplinary international team comprising physicians, specialist nurses, a psychologist, software engineers, academic researchers, cancer survivors, patient ambassadors, and ostomy care medical equipment suppliers was formed to facilitate the development of this patient-centric digital solution. The process occurred in several stages: (1) review of current practice

through clinic visits and on-site observations; (2) review of literature and findings from preliminary studies; (3) content development grounded in an underpinning theory; (4) integration of support services; and (5) optimizing user experience through improving interface aesthetics and customization. In our study, 5 participants with CRC performed preliminary assessments on the quality of the developed solution using the 20-item user version of the Mobile App Rating Scale (uMARS), which had good psychometric properties.

Results: Based on the collected uMARS data, the smartphone app was rated highly for functionality, aesthetics, information quality, and perceived impact, and moderately for engagement and subjective quality. Several limiting factors such as poor agility in the adoption of digital technology and low eHealth literacy were identified despite efforts to promote engagement and ensure ease of use of the mobile app. To overcome such barriers, additional app-training sessions, an instruction manual, and regular telephone calls will be incorporated into the iCanManage program during the trial period.

Conclusions: This form of multidisciplinary collaboration is advantageous as it can potentially streamline existing care paths and allow the delivery of more holistic care to the CRC population during the perioperative period. Should the program be found to be effective and sustainable, hospitals adopting this digital solution may achieve better resource allocation and reduce overall health care costs in the long run.

Trial Registration: ClinicalTrials.gov NCT04159363; https://clinicaltrials.gov/ct2/show/NCT04159363

(J Med Internet Res 2021;23(12):e31917) doi:10.2196/31917

KEYWORDS

colorectal cancer; digital solutions; mobile health; psychosocial; mHealth; smartphone app; mobile phone app

Introduction

Colorectal cancer (CRC) is known to be one of the leading causes of cancer-related morbidity and mortality among both men and women worldwide [1]. It will continue to be a global threat because of its profound impact on societies and public health [2]. Surgical tumor resection is the primary treatment for colorectal cancer, and as worldwide incidences continue to surge, the corresponding demand for elective surgeries is expected to rise as well [3]. The period leading up to and following an operation can be exceptionally sensitive and challenging for some patients, given the state of psychological vulnerability they are in, the debilitating symptoms, and the lifestyle changes they concurrently contend with [4]. Under the current enhanced recovery program or "fast-track surgery," much of the preoperative physiologic optimization occurs outside of hospital [5,6], and shortened hospitalization stays further mandate or pressurize CRC survivors to be independent and competent in self-care and disease management postdischarge [7,8]. Despite the introduction of various interventions in recent years, frequent, unwarranted hospital readmissions resulting from postoperative complications (eg, surgical site infections, intestinal obstruction, bleeding, and ostomy malfunction) imply that moderate self-efficacy levels reported among CRC survivors are suboptimal [9-12]. Furthermore, informal caregivers have been reported to lack confidence and have inadequate training in managing their care recipients' bowel problems, pain, fatigue, medications, and other symptoms [13]. These combined findings justify the need for more comprehensive and targeted measures. Therefore, this paper aimed to describe the development of a smartphone-based interactive CRC self-management psychosocial program (iCanManage) in order to improve perioperative health outcomes among patients with CRC undergoing elective CRC surgeries and their family caregivers.

Methods

Overview

The initial stage involved the establishment of a multidisciplinary international team by the study's principal investigator. Following the recruitment of a few academics, clinicians from the 2 largest acute hospitals in Singapore, ostomy care appliance specialists, as well as software engineers from a renowned Finnish company were approached. Meetings were organized to assemble all parties to brainstorm for ideas and kick-start the construction and mapping of timelines delineating critical milestones to be reached. The roles and the respective responsibilities were delegated (Table 1), and a reporting cascade was put in place to facilitate consistency and consensus throughout the development of the patient-centric digital solution. The appointed project manager assisted with the arrangement of all team-based activities and timely documentation of the project's progress.

Table 1. Multidisciplinary team members and roles.

Key personnel	Partner country	Role in the development process
Principal investigator	Singapore	• Conceptualize and supervise the iCanManage development and trial process
Project manager	Singapore	• Coordinate communication among collaboration partners and key stakeholders
Medical doctors	Singapore	• Provide content expertise for the creation of surgery care paths
Specialist nurses	Singapore	• Provide content expertise on perioperative assessments and counseling, symptom management, and self-care
Psychologist	Singapore	• Provide content expertise to develop, structure, and arrange mindfulness-based practices and activities • Provide formal psychological support
Software engineers	Finland and Singapore	• Design digital solution, input content within care path, and maximize functionality and customization of the software interface
Academic researchers	Singapore and Finland	• Collect field data, expert opinion, and public feedback before trial testing
Cancer survivors and patient ambassadors	Singapore	• Provide informal peer support through sharing of personal experiences, identification of common challenges, and coping strategies. • Provide information on external support services
Cancer institute and society	Singapore	• Provision of patient education pamphlets, brochures, and booklets
Ostomy care medical equipment suppliers	Denmark (Coloplast) and United States (ConvaTec)	• Provide multimedia on ostomy care, instructional guidelines, and practical tips on lifestyle adjustment after stoma creation

Review of Current Clinical Practice

To grasp a better understanding and to identify gaps in the current perioperative workflow of local hospital outpatient settings, the academic study team members conducted multiple clinic visits and on-site observations. Physical examinations, clinical procedures, care plan discussions, as well as types of queries and concerns raised by the patients and their accompanied caregivers during medical consultation sessions, were manually recorded. Preoperative assessments and instructions, educational reference materials, and relevant counseling advice provided by specialist nurses based on the patients' physiological status were also collected and documented. Besides the aforementioned, some study team members attended in-house peer support sessions, organized by patient ambassadors, and talks involving ostomy awareness to get a sense of the topics shared and activities availed for participation. These public events enabled conversations exploring patients with CRC and their caregivers' struggles, needs, satisfaction with care, and support preferences. Information about the referral process for formal psychosocial support was sought for knowledge about equity of access to such resources.

Through these reviewed processes, a few critical issues were identified. Apart from the occasional linguistic mismatch, a handful of patients, in particular those who attended the preoperative counseling alone, expressed confusion at the information provided by the specialist nurses. Caregivers who were unavailable for the face-to-face teaching were often left at a loss, resulting in multiple calls to the hospital for clarifications. Moreover, nurses themselves opined that the session was too overwhelming for those still fragile from receiving their diagnosis. They noticed that many of the seniors with CRC had difficulty retaining the given instructions and misplaced the reference materials provided. It was a constant challenge for 1 of the hospitals to ensure that the patients and their caregivers were competent with stoma care within the short hospitalization stay before discharge. Furthermore, peer-sharing sessions came to a halt as soon as COVID-19 emerged, further depriving the patients of the already minimal psychosocial support. These gaps highlighted major deficiencies in current practice and triggered this initiative to involve novel technologies to address the complex needs of this population.

Review of Literature and Findings From Preliminary Studies

The existing empirical evidence postulates that individuals with higher self-efficacy are more likely to be able to overcome illness-related challenges. Within the CRC population, substantial literature has reported positive associations between self-efficacy, adjustment to colostomy, lifestyle adaptation, symptom coping, social support, physical health, psychological well-being, and quality of life [8,14-19]. One study found presurgery self-efficacy, anxiety, and depression to be important predictors of one's recovery [12], and this is a significant finding because of the interdependent relationships between patients and their caregivers' self-efficacy as well as physical and mental health [20,21].

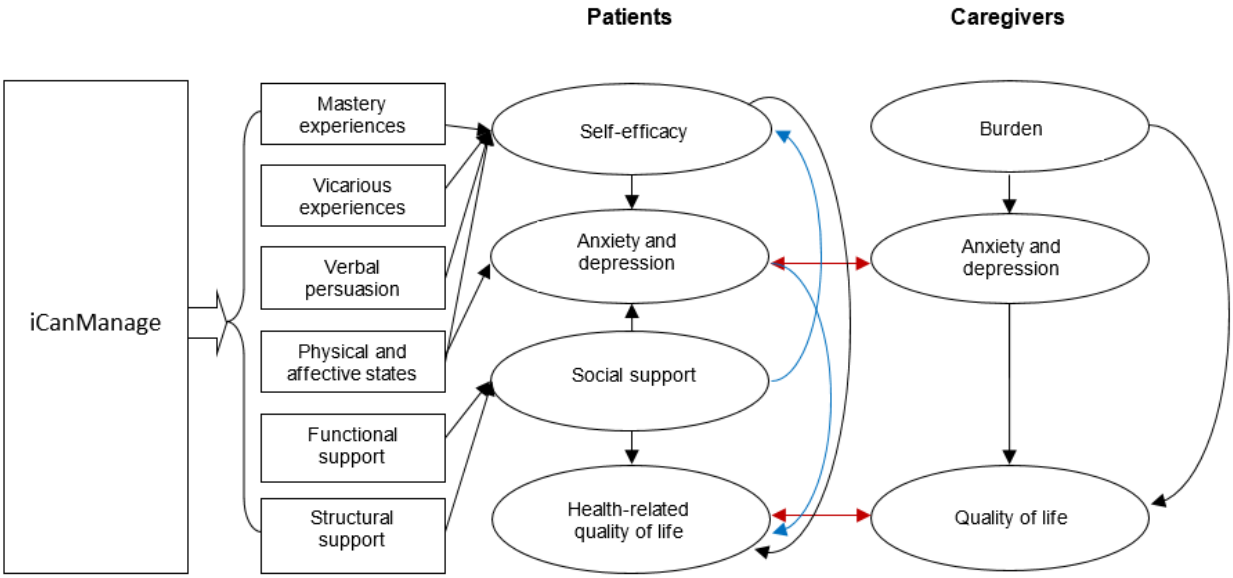
For CRC patients in particular, psychosocial care plays a subsidiary role in biomedical treatment where its positive effects on multiple mental health outcomes and quality of life have been verified in several reviews [22-24]. However, as

COVID-19 continues to persist, conventional face-to-face perioperative counseling by health care professionals has become increasingly challenging. Access to psychosocial support is limited and confined to those deemed to be at risk of significant psychological morbidity. This affects up to approximately 160 patients, who undergo elective colorectal resections annually in 1 local health care institution alone [25], and aggravates the mental distress already prevalent within this population [26]. Moreover, a preliminary study found great appetite for easily retrievable, in-depth, credible information about treatment and interest in the use of health technologies with embedded multimedia [27]. Such indications call for the need to review existing care delivery models and consider possible shifts toward the adoption of telemedicine.

Content Development Based on Self-efficacy Theory and Literature

Having put together all the gathered information, we developed the iCanManage program based on the self-efficacy theory by Bandura [28] and the interrelationships between self-efficacy, anxiety, depression, social support, burden, and quality of life, as previously described (Figure 1). This theory rides on the core assumption that the likelihood of a person engaging in a behavior depends on their conviction that they are able to successfully execute the course of action required to produce the intended outcome. Once the behavior has been adopted, the individual estimates and evaluates whether the given behavior leads to certain outcomes. In this theory, it is postulated that self-efficacy stems from four sources, namely mastery experiences, vicarious experiences, verbal persuasion, and emotional and physiological states, which we will describe in the following section.

Figure 1. Conceptual framework for the iCanManage program.



Mastery Experiences

Mastery experiences are especially influential, and they accumulate from persistent and sustained efforts in achieving performance success and overcoming failures, obstacles, and adversity. The self-efficacy developed through mastery experiences refers to resilience and the ability to quickly rebound when faced with setbacks.

Vicarious Experiences

Vicarious experiences are gained from assimilation to the successes and failures of social models. Through the observation of similar individuals who succeed, people come to believe that they too are capable of succeeding. Conversely, observing similar individuals who fail despite their high effort will lower one's expectations and prevent self-criticism.

Verbal Persuasion

Verbal persuasion from social interactions can also enhance one's perceived self-efficacy. Through the use of positive words, people's belief that they have the capacity to succeed is strengthened, and they are motivated to mobilize greater effort to sustain or confront the challenges that come their way.

Physical and Affective States

Self-efficacy is also influenced by how an individual views and understands emotional and physical reactions, including stress, tension, fatigue, and mood states, to be a reflection of their performance abilities. By learning how to regulate their responses when encountering difficult and challenging situations, one can cultivate a greater locus of control and develop self-efficacy.

CRC surgeries, unlike those for other cancer types, bring about many physical and lifestyle changes that require the acquisition of new behavior patterns [29]. Yet, successful adaptation and adjustment usually follow a feedback loop involving response consequences and predictive cues, which are largely influenced by the individual's cognitive processes and perceived locus of control [28]. Bandura's self-efficacy theory has been widely used within the cancer population to guide intervention

development, where significant positive effects on numerous health outcomes were reported [19,30]. Knowing how demanding and unpredictable the course of this illness is, it became imperative to design the iCanManage program to empower our trial participants by enabling access to the four sources of self-efficacy. By doing so, the participants will have greater preparedness for surgery, which may benefit recovery after surgery [31]. At the same time, they will also be better positioned to respond and withstand the ordeal in the long run [14]. A description of how the contents were designed based on this theory is presented in Multimedia Appendix 1.

Integration of Support Services

Apart from formal professional support, tangible and practical help are also essential for sustained adaptive coping, especially after an individual is discharged from the health care institution and reintegrates back into the community [4,32]. Hence, cancer survivors and patient ambassadors of external support groups (eg, Ostomy Association of Singapore, Singapore Cancer Society, and Patient Advocacy Network) were invited to share their personal experiences in the form of videos to relay useful messages on resuming activities of daily living while at the same time providing companionship throughout the receiving patient's cancer journey. These videos were later embedded within the digital solution.

In Singapore, a majority of patients with CRC fitted with a stoma (ie, ileostomy or colostomy) obtain their ostomy care appliances from Coloplast (Coloplast A/S) and ConvaTec (ConvaTec Group plc). For many years, these 2 companies from Denmark and the United States, respectively, have been committed to providing advanced therapies for wound, ostomy, and continence care [33,34]. To ensure our trial participants are able to manage their stoma and quickly familiarize themselves with the use of different ostomy products, partnerships with these companies were built to borrow educational multimedia for dissemination through the digital solution. This collaboration also opened a platform for the trial participants to connect with the product suppliers and ostomy care specialists directly.

Optimizing User Experience

Once the final version of the proposed contents was ready, software engineers from Buddy Healthcare were engaged to input all materials into the created digital solution known as BuddyCare. Buddy Healthcare is known for its pioneering efforts in designing award-winning automated care pathways for improved care coordination and patient engagement [35]. To date, they have over 200 care pathway templates for elective surgeries in use by hospitals across various continents (Europe, United States, and Asia Pacific), where positive results have been attained, including significant reductions in preoperative phone calls and the time spent on counseling, as well as high recommendation ratings from health care professionals who have subscribed to the platform.

To ensure seamless navigation and optimal user experience, iterative processes of trial and error took place to format and arrange the content layout to be displayed. During these evaluation sessions, some notable points were raised: (1) ensuring a sizeable but digestible amount of information load; (2) modulating the appropriate time spent and the daily app usage; (3) using elderly friendly fonts, text size, and colors; (4) using icons to increase the visibility of new instructions; and (5) controlling the frequency of reminders, alerts, and notifications.

After the smartphone-based digital solution was fine-tuned, it was formally implemented as part of the intervention component of a multicenter randomized controlled trial (NCT04159363) conducted at 2 large-scale acute tertiary hospitals. At present, the full-scale randomized controlled trial is ongoing, and the data are insufficient for analysis. However, apart from our main study outcomes, preliminary assessments and feedback on BuddyCare were solicited using the Mobile App Rating Scale: User Version (uMARS). This 20-item instrument was developed by Stoyanov and colleagues to measure the quality of mHealth apps across the 5 subscales of engagement (5 items), functionality (4 items), aesthetics (3 items), information quality (4 items), and the mHealth app's subjective quality (4 items) [36]. Each question is rated against 5 different response options, which are later summed and averaged to generate a mean score for the individual subscales. Mean scores range from 1 to 5, with higher scores denoting better quality in the particular domain represented by the subscale. This tool has been used to evaluate mHealth apps on well-being [37] and mindfulness [38], and it has been found to possess high internal consistency (Cronbach α=.90) and test-retest reliability (intraclass correlation coefficient range 0.66-0.70). Additionally, it is pitched at a readability level of those aged 12 to 13 years, which also renders it suitable for use among elderly participants who are less educated.

Prior to study initiation, the team sought and received approval from the local ethics board (National Healthcare Group Domain-Specific Review Board [2019/01002]). Written informed consent was obtained from all enrolled participants, and the respective institutions' personal data protection act will be strictly adhered to during the trial period.

Results

The iCanManage Mobile App

Our eventual product was the iCanManage mobile app on the platform of BuddyCare, which has three special features: a surgical timeline, search functionality, and a menu tab. The timeline stretches over a 29-day perioperative phase (14 days before surgery, the day of surgery, and 14 days after surgery) where users will receive information packages listing important tasks about how to prepare for surgery, postsurgery monitoring, and discharge care on a daily basis. In addition, users will also be introduced to mindfulness-based practices, which are purposed to facilitate emotional coping and regulation. Examples of such include mindful breathing, body scan, loving-kindness meditation, and mindful coping. Daily tasks are color coded based on the urgency for review and completion, and reminders are sent at scheduled timings to ensure tasks are acknowledged before they become overdue. Positive slogans are also refreshed periodically to motivate users to maintain a positive outlook or comfort them in their setbacks.

The search functionality enables participants to locate and retrieve information within the timeline easily by entering keywords. This is necessary given the vast amount of text, files, and multimedia materials embedded in the app, which may only become relevant for the user at specific time points of the treatment continuum. Users may also revisit information retrospectively or prospectively as they wish.

Under the menu tab, users will find peer support videos, ostomy care videos, and help hotlines. Through the messaging feature, users may also communicate their concerns to the attending care team and seek tailored advice. Pictures of wound and stoma conditions can be sent anonymously, and remote surveillance and consultation can be conducted. Through this digital solution, we hope users will be more equipped, confident, and independent when undergoing elective CRC surgery. During the trial period, the participants randomized to the intervention arm will be given an activation code that permits personalized access to the digital care path within the mobile app installed on their handheld devices. By doing so, individuals may feel more invested in their own treatment progress and work toward their set goals. An overview of the components and content is outlined in Multimedia Appendices 2 and 3, while screenshots of the developed solution are shown in Multimedia Appendix 4 (image A: English version; image B: Chinese version).

Preliminary Findings on the Quality of BuddyCare

The uMARS survey data from 5 patients with CRC were analyzed descriptively. Among these participants, 4 were females and 1 was male. The mean age of these participants was 57.8 years old (SD 3.77) (range: 54 to 64 years old). The mean subscale scores were 2.8/5 (SD 0.40) for engagement, 3.9/5 (SD 0.34) for functionality, 3.5/5 (SD 0.38) for aesthetics, 3.4/5 (SD 0.34) for information quality, and 2.4/5 (SD 0.42) for app subjective quality. Besides these 5 subscales, an additional component of the uMARS known as perceived impact (consisting of 6 items) rated from “1=strongly disagree” to “5=strongly agree” was measured. The mean score for perceived impact was 4 out of 5, and the overall app quality mean score was tabulated to be 3.4 out of 5. Moreover, 2 participants provided additional comments in the empty field of the questionnaire; 1 of them stated that some of the contents within the app were repetitive, while the other said it was good to try out a new program, and that it was a well-done initiative with room for improvement.

Strategies to Overcome Potential Limitations

Although much has been done to increase the usability and acceptability of the digital solution, there remain several potential limitations. For instance, elderly users who are not technology savvy may encounter difficulties when navigating the mobile app and, as a result, become more confused and distressed. Once a scenario as such occurs, users tend to become hesitant or resistant to subsequent health technologies. Moreover, poor eHealth literacy may further hinder their understanding of the presented information. Given the high incidence rates proportionate to the age of individuals diagnosed with CRC in the local context [39], we propose the inclusion of family caregivers during the trial to avoid underutilization of the app and its benefits being diminished. Study team members will provide face-to-face training on the day of recruitment to introduce all features, functions, and materials within the app. An instruction manual will be provided at the end of the training session for reference should any technical issues render the user unable to return to the app’s main page. Moreover, telephone calls will be made by academic researchers weekly to ensure the app is operating properly. Lastly, all contents were translated and will be made available in simplified Chinese language to suit the linguistic demographics of the local population. That being said, our multicenter trial will recruit both English- and Chinese-speaking patients scheduled to undergo elective CRC surgeries.

Discussion

Principal Findings

Telemedicine has been around for some time, but it was not until lately that a growing interest in interactive health communication technologies has been observed among users of health care and their providers [40]. As the world continues to battle COVID-19 and countries strive to embrace smart technologies, there is a compelling need for more initiatives to consistently lead seniors toward the adoption of digitalization, and our trial paved the way for such opportunities. In this regard, our multidisciplinary international collaboration also unveiled the following valuable insights.

Firstly, as with most studies, researchers are inclined to focus only on raising the technical skill level of the users, often forgetting that health care providers themselves ought to be knowledgeable and proficient in the said technology [41]. Cancer care is complex and requires multimodal treatment over prolonged periods of time. As mentioned earlier, disease management and self-care for patients with CRC can be exceptionally trying due to altered bowel patterns, dietary intolerance, and stoma-related concerns [42]. Multiple points of contact with various allied health care professionals are common and necessary, yet this would mean frequent hospital visits. Our collaboration is hopeful that health technologies will strengthen the delivery of holistic care through better integration of medical data, standardization of educational counseling, and increased patient autonomy. To achieve this, barriers prohibiting health care professionals from actual utilization, such as shortage of easily accessible and high connectivity computers, security issues, and the lack of staff training, should be eliminated [43].

Secondly, much controversy has been associated with web-based or online psychotherapies. The use of technology is generally perceived to impede the interpersonal therapeutic relationship between the psychologist and the patient, thereby reducing treatment effectiveness [44]. However, input from our multidisciplinary teammates ushered possibilities of transforming traditional, formal psychotherapy into informal practices that can be applied or exercised in the individual’s daily life. While such options offer greater flexibility [45], careful attention should be paid to maintain open channels for communication and assessment of patients’ responses, as well as to safeguard the confidentiality of information shared over the internet platform being used, which will be further explored in our trial.

Lastly, preliminary uMARS data from the actualization of our digital solution raised several points for reflection and future consideration. Although these findings cannot be generalized to the entire CRC population, scores of each subscale provided a glimpse of the participants' preferences and helped gauge how well the digital solution addressed their needs. According to the results, high levels of functionality, aesthetics, information quality, and perceived impact indicated that the BuddyCare mobile app is generally usable, pleasant to look at, credible, and relevant for our participants' medical condition. Conversely, moderate levels of engagement revealed that our digital solution was still lacking in interactivity and customization. This could mean that participants desire an exchange of information between themselves and the software, researchers, or clinicians, even though the enablement of such feature may increase the technical complexity of the app. In terms of subjective quality, lower-than-midpoint values suggest that the app was likely not convincing enough to warrant a recommendation to others or payment for its usage in the long run. This is reasonable given the relatively young age of these 5 participants who reported the data. The young, being more exposed to digital technologies, may have higher expectations when evaluating new mobile apps. Moreover, our findings agree with some of the consolidated themes derived from stakeholder perceptions delineated in a study by Mercer and colleagues [46]. These include the need to explore flexibility in design to achieve sustained usage, as well as to leverage the knowledge of behavioral psychology to improve user engagement. Considering these pertinent issues, more data from the full-scale randomized controlled trial will be necessary to verify and support these findings before any change is to be made to the digital solution.

Conclusion

This paper detailed the sequential development of the iCanManage smartphone-based psychosocial program to be used in a multicenter randomized controlled trial. This patient-centric digital solution will be the first to deliver a combination of surgery and psychosocial care to patients scheduled to undergo elective CRC surgeries and their family caregivers. If found to be effective, the iCanManage will see increased adherence to prescribed treatments, increased patient satisfaction through better health outcomes, and smoother transitions from hospital to home following surgery. In the long run, better allocation of personnel, time, and costs can be attained, and the economic burden can be significantly reduced when extended to elective surgeries for other disease types.

Acknowledgments

The authors wish to thank the clinicians, administrators, and those who have directly and indirectly contributed their expert or public opinion to the development of this program. This study would not have been possible without the support of the National University of Singapore, Singapore General Hospital, National University Hospital in Singapore, University of Oulu, and Buddy Healthcare Ltd Oy in Finland. The team deeply appreciates the funding provided by the Singapore Cancer Society and Business Finland. This research work is funded by the Singapore Cancer Society Cancer Research Grant 2018 and the Business Finland Grant 203/31/2018.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theory-guided contents.
[DOCX File , 17 KB - jmir_v23i12e31917_app1.docx]

Multimedia Appendix 2
Overview of the components and contents within the iCanManage mobile app.
[DOCX File , 27 KB - jmir_v23i12e31917_app2.docx]

Multimedia Appendix 3
Daily outline of contents.
[DOCX File , 18 KB - jmir_v23i12e31917_app3.docx]

Multimedia Appendix 4
Screenshot images of the completed digital solution in English and Chinese versions.
[DOCX File , 924 KB - jmir_v23i12e31917_app4.docx]

References

1. Douaiher J, Ravipati A, Grams B, Chowdhury S, Alatise O, Are C. Colorectal cancer-global burden, trends, and geographical variations. J. Surg. Oncol 2017 Feb 13;115(5):619-630. [doi: 10.1002/jso.24578]

2. Araghi M, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, et al. Global trends in colorectal cancer mortality: projections to the year 2035. Int J Cancer 2019 Jun 15;144(12):2992-3000 [FREE Full text] [doi: 10.1002/ijc.32055] [Medline: 30536395]
3. Rentsch M, Schiergens T, Khandoga A, Werner J. Surgery for Colorectal Cancer - Trends, Developments, and Future Perspectives. Visc Med 2016 Jun 13;32(3):184-191 [FREE Full text] [doi: 10.1159/000446490] [Medline: 27493946]
4. Hildebrandt C, Mayer H, Koller A. Experiences of patients with colorectal cancer from diagnosis until completion of treatment: A meta-ethnography approach. Psychooncology 2019 Feb 28;28(2):219-227. [doi: 10.1002/pon.4946] [Medline: 30447170]
5. Gustafsson UO, Scott MJ, Hubner M, Nygren J, Demartines N, Francis N, et al. Guidelines for Perioperative Care in Elective Colorectal Surgery: Enhanced Recovery After Surgery (ERAS) Society Recommendations: 2018. World J Surg 2019 Mar 13;43(3):659-695. [doi: 10.1007/s00268-018-4844-y] [Medline: 30426190]
6. Gillis C, Gill M, Marlett N, MacKean G, GermAnn K, Gilmour L, et al. Patients as partners in Enhanced Recovery After Surgery: A qualitative patient-led study. BMJ Open 2017 Jun 24;7(6):e017002 [FREE Full text] [doi: 10.1136/bmjopen-2017-017002] [Medline: 28647727]
7. Johansson A, Brink E, Cliffordson C, Axelsson M. The function of fatigue and illness perceptions as mediators between self-efficacy and health-related quality of life during the first year after surgery in persons treated for colorectal cancer. J Clin Nurs 2018 Apr 26;27(7-8):e1537-e1548. [doi: 10.1111/jocn.14300] [Medline: 29399917]
8. Qian H, Yuan C. Factors associated with self-care self-efficacy among gastric and colorectal cancer patients. Cancer Nurs 2012;35(3):E22-E31. [doi: 10.1097/NCC.0b013e31822d7537] [Medline: 22067688]
9. Slankamenac K, Slankamenac M, Schlegel A, Nocito A, Rickenbacher A, Clavien P, et al. Impact of postoperative complications on readmission and long-term survival in patients following surgery for colorectal cancer. Int J Colorectal Dis 2017 Jun 14;32(6):805-811. [doi: 10.1007/s00384-017-2811-y] [Medline: 28411352]
10. Sheetz K, Waits S, Krell R, Morris A, Englesbe M, Mullard A. Complication rates of ostomy surgery are high and vary significantly between hospitals. Diseases of the colon and rectum 2014;57(5):632-637. [doi: 10.1097/dcr.0000000000000038]
11. Butler D. Early postoperative complications following ostomy surgery: a review. J Wound Ostomy Continence Nurs 2009;36(5):513-519. [doi: 10.1097/WON.0b013e3181b35eaa] [Medline: 19752661]
12. Foster C, Haviland J, Winter J, Grimmett C, Chivers Seymour K, Batehup L, Members of the Study Advisory Committee. Pre-Surgery Depression and Confidence to Manage Problems Predict Recovery Trajectories of Health and Wellbeing in the First Two Years following Colorectal Cancer: Results from the CREW Cohort Study. PLoS One 2016 May 12;11(5):e0155434 [FREE Full text] [doi: 10.1371/journal.pone.0155434] [Medline: 27171174]
13. Havyer RD, van Ryn M, Wilson PM, Griffin JM. The effect of routine training on the self-efficacy of informal caregivers of colorectal cancer patients. Support Care Cancer 2017 Apr 26;25(4):1071-1077. [doi: 10.1007/s00520-016-3494-6] [Medline: 27889827]
14. Grimmett C, Haviland J, Winter J, Calman L, Din A, Richardson A, et al. Colorectal cancer patient's self-efficacy for managing illness-related problems in the first 2 years after diagnosis, results from the ColoREctal Well-being (CREW) study. J Cancer Surviv 2017 Oct 19;11(5):634-642 [FREE Full text] [doi: 10.1007/s11764-017-0636-x] [Medline: 28822053]
15. Wu H, Chau J, Twinn S. Self-efficacy and quality of life among stoma patients in Hong Kong. Cancer Nurs 2007;30(3):186-193. [doi: 10.1097/01.ncc.0000270704.34296.86]
16. Jin Y, Ma H, Jiménez-Herrera M. Self-disgust and stigma both mediate the relationship between stoma acceptance and stoma care self-efficacy. J Adv Nurs 2020 Oct 23;76(10):2547-2558. [doi: 10.1111/jan.14457] [Medline: 32700799]
17. Lim SH, Chan SWC, Chow A, Zhu L, Lai JH, He H. Pilot trial of a STOMA psychosocial intervention programme for colorectal cancer patients with stomas. J Adv Nurs 2019 Jun 21;75(6):1338-1346. [doi: 10.1111/jan.13973] [Medline: 30740765]
18. Anderson AS, Steele R, Coyle J. Lifestyle issues for colorectal cancer survivors--perceived needs, beliefs and opportunities. Support Care Cancer 2013 Jan 8;21(1):35-42. [doi: 10.1007/s00520-012-1487-7] [Medline: 22773297]
19. Zhang M, Chan SW, You L, Wen Y, Peng L, Liu W, et al. The effectiveness of a self-efficacy-enhancing intervention for Chinese patients with colorectal cancer: a randomized controlled trial with 6-month follow up. Int J Nurs Stud 2014 Aug;51(8):1083-1092. [doi: 10.1016/j.ijnurstu.2013.12.005] [Medline: 24406033]
20. Kershaw T, Ellis KR, Yoon H, Schafenacker A, Katapodi M, Northouse L. The Interdependence of Advanced Cancer Patients' and Their Family Caregivers' Mental Health, Physical Health, and Self-Efficacy over Time. Ann Behav Med 2015 Dec 21;49(6):901-911 [FREE Full text] [doi: 10.1007/s12160-015-9743-y] [Medline: 26489843]
21. Kayser K, Acquati C, Reese JB, Mark K, Wittmann D, Karam E. A systematic review of dyadic studies examining relationship quality in couples facing colorectal cancer together. Psychooncology 2018 Jan 26;27(1):13-21. [doi: 10.1002/pon.4339] [Medline: 27943551]
22. Mosher CE, Winger JG, Given BA, Shahda S, Helft PR. A systematic review of psychosocial interventions for colorectal cancer patients. Support Care Cancer 2017 Jul 22;25(7):2349-2362 [FREE Full text] [doi: 10.1007/s00520-017-3693-9] [Medline: 28434094]
23. Hoon LS, Chi Sally CW, Hong-Gu H. Effect of psychosocial interventions on outcomes of patients with colorectal cancer: a review of the literature. Eur J Oncol Nurs 2013 Dec;17(6):883-891. [doi: 10.1016/j.ejon.2013.05.001] [Medline: 23759360]

24. Son H, Son Y, Kim H, Lee Y. Effect of psychosocial interventions on the quality of life of patients with colorectal cancer: a systematic review and meta-analysis. Health Qual Life Outcomes 2018 Jun 08;16(1):119 [FREE Full text] [doi: 10.1186/s12955-018-0943-6] [Medline: 29884182]
25. Seow-En I, Tan WJ, Dorajoo SR, Soh SHL, Law YC, Park SY, et al. Prediction of overall survival following colorectal cancer surgery in elderly patients. World J Gastrointest Surg 2019 May 27;11(5):247-260 [FREE Full text] [doi: 10.4240/wjgs.v11.i5.247] [Medline: 31171956]
26. Peng Y, Huang M, Kao C. Prevalence of Depression and Anxiety in Colorectal Cancer Patients: A Literature Review. Int J Environ Res Public Health 2019 Jan 31;16(3):411 [FREE Full text] [doi: 10.3390/ijerph16030411] [Medline: 30709020]
27. Lim SH, Chan SWC, Lai JH, He H. A qualitative evaluation of the STOMA psychosocial intervention programme for colorectal cancer patients with stoma. J Adv Nurs 2019 Jan 17;75(1):108-118. [doi: 10.1111/jan.13821] [Medline: 30109721]
28. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review 1977;84(2):191-215. [doi: 10.1037/0033-295x.84.2.191]
29. Nikoletti S, Young J, Levitt M, King M, Chidlow C, Hollingsworth S. Bowel problems, self-care practices, and information needs of colorectal cancer survivors at 6 to 24 months after sphincter-saving surgery. Cancer Nurs 2008;31(5):389-398. [doi: 10.1097/01.ncc.0000305759.04357.1b]
30. Merluzzi TV, Pustejovsky JE, Philip EJ, Sohl SJ, Berendsen M, Salsman JM. Interventions to enhance self-efficacy in cancer patients: A meta-analysis of randomized controlled trials. Psychooncology 2019 Sep 09;28(9):1781-1790 [FREE Full text] [doi: 10.1002/pon.5148] [Medline: 31206917]
31. Öhlén J, Sawatzky R, Pettersson M, Sarenmalm EK, Larsdotter C, Smith F, et al. Preparedness for colorectal cancer surgery and recovery through a person-centred information and communication intervention - A quasi-experimental longitudinal design. PLoS One 2019 Dec 12;14(12):e0225816 [FREE Full text] [doi: 10.1371/journal.pone.0225816] [Medline: 31830066]
32. Abelson JS, Chait A, Shen MJ, Charlson M, Dickerman A, Yeo H. Coping strategies among colorectal cancer patients undergoing surgery and the role of the surgeon in mitigating distress: A qualitative study. Surgery 2019 Feb;165(2):461-468. [doi: 10.1016/j.surg.2018.06.005] [Medline: 30316575]
33. Improving the lives of people we touch 2021. ConvaTec. 2021. URL: https://www.convatec.com/# [accessed 2021-06-28]
34. Coloplast care can make a difference for you. Coloplast. URL: https://www.coloplast.sg/ [accessed 2021-06-28]
35. We automate care pathways. Buddy Healthcare. URL: https://www.buddyhealthcare.com/en/ [accessed 2121-06-28]
36. 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 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
37. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422]
38. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
39. Singapore Cancer Registry Annual Report 2018. National Registry of Diseases Office. 2021 Mar 31. URL: https://www.nrdo.gov.sg/docs/librariesprovider3/default-document-library/scr-annual-report-2018.pdf?sfvrsn=bcf56c25_0 [accessed 2021-11-19]
40. Schmitz-Grosz K. Changes in Medical Processes Due to Digitalization: Examples from Telemedicine. Digitalization in Healthcare 2021:73-92. [doi: 10.1007/978-3-030-65896-0_7]
41. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. The Lancet Oncology 2020 May;21(5):e240-e251. [doi: 10.1016/s1470-2045(20)30021-8]
42. Smith F, Öhlén J, Persson L, Carlsson E. Daily Assessment of Stressful events and Coping in early post-operative recovery after colorectal cancer surgery. Eur J Cancer Care (Engl) 2018 Mar 24;27(2):e12829. [doi: 10.1111/ecc.12829] [Medline: 29363821]
43. Whitten P, Notman M, Maynard C, Henry R, Glandon R. Interactive Health Communication Technologies in the Public Health Sector: Positive Perceptions Still Outpace Actual Utilization. Journal of Technology in Human Services 2004 Jun 07;22(3):25-40. [doi: 10.1300/j017v22n03_02]
44. Baker KD, Ray M. Online counseling: The good, the bad, and the possibilities. Counselling Psychology Quarterly 2011 Dec;24(4):341-346. [doi: 10.1080/09515070.2011.632875]
45. Barak A, Grohol JM. Current and Future Trends in Internet-Supported Mental Health Interventions. Journal of Technology in Human Services 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
46. Mercer K, Baskerville N, Burns CM, Chang F, Giangregorio L, Tomasson Goodwin J, et al. Using a collaborative research approach to develop an interdisciplinary research agenda for the study of mobile health interventions for older adults. JMIR Mhealth Uhealth 2015 Feb 10;3(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3509] [Medline: 25669321]

Abbreviations

CRC: colorectal cancer
uMARS: Mobile App Rating Scale, user version

Edited by R Kukafka; submitted 10.07.21; peer-reviewed by J Zhu, S Azadnajafabad; comments to author 25.08.21; revised version received 10.09.21; accepted 21.10.21; published 07.12.21.

Please cite as:
Wan SW, Chong CS, Toh EL, Lim SH, Loi CTT, Lew YFH, Chua MCH, Jee XP, Liu G, Zhu L, Pikkarainen M, He HG
A Theory-Based, Multidisciplinary Approach to Cocreate a Patient-Centric Digital Solution to Enhance Perioperative Health Outcomes Among Colorectal Cancer Patients and Their Family Caregivers: Development and Evaluation Study
J Med Internet Res 2021;23(12):e31917
URL: https://www.jmir.org/2021/12/e31917
doi:10.2196/31917
PMID:34878991

Original Paper

Internet and Face-to-face Cognitive Behavioral Therapy for Postnatal Depression Compared With Treatment as Usual: Randomized Controlled Trial of MumMoodBooster

Jeannette Milgrom[1,2], PhD; Brian G Danaher[3,4], PhD; John R Seeley[3,4], PhD; Christopher J Holt[5], PhD; Charlene Holt[1], PsyD; Jennifer Ericksen[1], MPsych; Milagra S Tyler[3,4], MA; Jeff M Gau[3,4], MS; Alan W Gemmill[1], PhD

[1]Parent-Infant Research Institute, Heidelberg Repatriation Hospital, Heidelberg Heights, Australia

[2]Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Australia

[3]University of Oregon, Eugene, OR, United States

[4]Influents Innovations, Eugene, OR, United States

[5]Australian College of Applied Psychology, Melbourne, Australia

Corresponding Author:
Alan W Gemmill, PhD
Parent-Infant Research Institute
Heidelberg Repatriation Hospital
300 Waterdale Rd
Heidelberg Heights, 3081
Australia
Phone: 61 3 9496 4496
Email: alan.gemmill@austin.org.au

Abstract

Background: Previous research has confirmed that symptoms of postnatal depression (PND) can be ameliorated through internet-delivered psychological interventions. Advantages of internet-delivered treatment include anonymity, convenience, and catering to women who are unable to access face-to-face (FTF) treatments. To date, no research has examined the efficacy of such interventions compared directly with FTF treatments in women clinically diagnosed with PND.

Objective: This study aims to compare the efficacy of one of the first web-based cognitive behavioral therapy (CBT) interventions (internet CBT+coach calls) for PND (*MumMoodBooster* [MMB]) with FTF-CBT in a randomized controlled trial (RCT).

Methods: In this study, 116 postnatal women with a *Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition* (*DSM-IV*) diagnosis of major or minor depression were randomized to MMB (39/116, 33.6%), FTF-CBT (39/116, 33.6%), or a treatment-as-usual (TAU) control condition (38/116, 32.8%). Diagnostic status was determined at baseline and at 21-week follow-up using the Structured Clinical Interview for the DSM-IV. Severity of anxiety and depressive symptoms was evaluated using the Depression Anxiety Stress Scales and the revised Beck Depression Inventory at baseline, 12-week follow-up (after treatment), and 21-week follow-up.

Results: Of the 116 participants, 107 (92.2%) had a diagnosis of major depression at baseline. Rates of remission from a major or minor depressive episode at 21 weeks in both the FTF-CBT and MMB groups were superior to that of the TAU group (56.6% and 47.7% less likely to be depressed, respectively) and they were not significantly different from each other. Although remission rates differed between TAU and FTF-CBT, growth models showed that, in terms of symptom reduction across time, the FTF-CBT treatment was not significantly better than TAU. By comparison, MMB was statistically superior to both TAU and FTF-CBT in reducing symptoms of depression, anxiety, and stress from baseline to the 21-week follow-up (large and moderate effect sizes). Thus, after 21 weeks, the average symptom scores for depression and anxiety of women receiving MMB were approximately half those of women in both the TAU and FTF-CBT groups.

Conclusions: In this RCT, MMB was at least as effective as FTF-CBT in achieving remission from a diagnosed PND episode. MMB was superior to TAU and FTF-CBT in encouraging and maintaining reduction of symptom severity over the 21-week follow-up for depressed postnatal women. These findings replicate results of prior studies on MMB that showed clinically significant improvements in depressive symptoms, and they provide direct empirical support that internet-delivered treatment for depressed postnatal women is a viable alternative to FTF treatment. The generalizability of the results needs to be examined

in future research, as RCTs of internet-based versus FTF treatments necessarily involve a subset of people who are willing to undertake either modality of treatment.

Trial Registration: Australia and New Zealand Clinical Trials Registry (ANZCTR) ACTRN12613000881730; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=364683&isReview=true

(J Med Internet Res 2021;23(12):e17185) doi:10.2196/17185

KEYWORDS

postnatal depression; postpartum depression; postnatal anxiety; postpartum anxiety; cognitive behavioral therapy; internet intervention; web-based intervention; randomized controlled trial; online intervention; treatment; mobile phone

Introduction

Background

A substantial proportion of new mothers suffer from postnatal depression (PND). A meta-analysis by Gavin et al [1] estimated a point prevalence of 12.9% at 3 months postpartum. Left untreated, PND has significant deleterious effects on the mother (eg, absence of well-being, feelings of lack of competence in the mothering role, and interactional difficulties with her infant), her partner (eg, partner's mental health and marital functioning or couple relationship problems), and her child (eg, diminished cognitive and psychosocial development) [2-5]. Children of mothers with PND are at an elevated risk of mental health difficulties during adolescence [6].

A considerable body of research supports the effectiveness of cognitive behavioral therapy (CBT) interventions for depression in clinical settings [7,8]. Approximately two-thirds of individuals in CBT trials are no longer diagnosed with depressive disorders at follow-up [8], and CBT also reduces the risk of relapse [9], particularly in cases of mild-to-moderate depression [10]. A range of treatment approaches is effective specifically for PND, and these have been systematically reviewed by Dennis and Hodnet [11] and Cuijpers et al [12]. Treatment approaches supported by positive research findings include CBT [13], counseling [14], pharmacotherapy [15], and interpersonal psychotherapy [16]. Milgrom et al [13] successfully adapted CBT treatments to the needs of new mothers with PND. These treatments were developed to be sensitive to specific needs during the perinatal period, for example, less time commitment to homework, techniques of relaxation on the run, and inclusion of women's partners where appropriate [13].

Reviews of the available evidence [17-19] have concluded that self-guided internet interventions are beneficial to depressed individuals but that the largest effect sizes are achieved when these are combined with guided human support, typically delivered via telephone or email contact. Such support can increase adherence to internet-delivered mental health treatments [20,21] and may also provide a *safety net* for individuals who may need additional help should their situation worsen. Alliance with a *trustworthy* coach is central to this model, and a number of studies have shown that a strong web-based working alliance can be achieved in guided internet-delivered treatments for posttraumatic stress disorder and depression [22,23]. The emerging picture is that very encouraging therapeutic effects can be achieved through structured internet programs, supported by low-intensity human guidance (typically <3 contact hours in a 6-week program) [17-19].

Effectiveness of Internet Psychological Interventions

The balance of existing evidence [18,24-26] suggests that psychological treatments delivered via the internet can be as effective as traditional face-to-face (FTF) approaches [27,28]. Moderate-to-large effects are generally reported [27], although it should be noted that, in many studies, the typical counterfactuals are waitlist controls and treatment as usual (TAU). For depression and anxiety in particular, the efficacy of internet-based treatment has been demonstrated relative to control conditions in populations with elevated symptoms and increasingly in clinically diagnosed groups [25,29-34]. Multiple meta-analyses and systematic reviews of internet interventions have confirmed clinically and statistically significant improvements in symptoms compared with TAU [25,27,34-36].

The Need for Internet Interventions Specific to PND

PND is undertreated: fewer than half of perinatal women seek help even when identified as depressed [37]. Barriers to clinic-based treatment uptake include fear of stigma, being perceived as a *bad* mother, feelings of failure, poor understanding of depression or the available help, concern about medication being passed through breastmilk [38], and difficulty attending because of the demands of the new baby [39]. Internet psychological interventions may offer advantages in this population. The perception that the internet is somewhat anonymous may assist in circumventing stigma [40]. Similarly, better accessibility may be achieved with a treatment program available from home. The flexibility of being able to review program content at any time on any day in a convenient location may help in minimizing childcare logistics and fitting in with changing and sleep-disrupted family schedules. Such interventions could then potentially provide psychological treatment to many women with PND who otherwise would not access support.

Spurred by these possibilities, a number of such internet interventions have been developed to address maternal perinatal depression. Beginning with our own work [41], the delivery of cognitive and behavioral internet interventions for perinatal symptoms of depression has emerged as a major focus. Several feasibility studies and randomized trials using CBT and behavioral activation have shown significant effects on postnatal depressive and anxiety symptoms, including those in the clinically severe range [42-46]. Compared with a TAU comparison group, Milgrom et al [43] reported a 4-fold increase in remission from a diagnosed depressive disorder using the

Structured Clinical Interview for the DSM-IV (SCID-IV) at 3 months after enrollment in the 6-session internet CBT intervention named MumMoodBooster (MMB; the US program version is named MomMoodBooster). O'Mahen et al [44,45,47] reported similar reductions in depressive symptoms with a 12-session intensive coaching behavioral activation therapy (the Net Mums program). Laughnan et al [48,49] have shown encouraging short-term effects for their brief, transdiagnostic internet intervention (MUMentum postnatal) that targets symptoms of both depression and anxiety. These internet interventions vary considerably in length and in the amount and nature of guidance or support offered (ranging from extended contacts with professional therapists to brief, low-intensity *technician* or coach support). The emerging picture suggests that internet interventions may help overcome several important barriers to seeking treatment in perinatal populations, and they have the potential to offer a scalable alternative for the delivery of psychological treatments.

Meta-analyses suggest that internet interventions have achieved medium-sized effects in reducing perinatal depressive symptoms, usually in comparison with TAU or waitlist control conditions [50-52]. To date, no published study has examined the efficacy of an internet intervention versus FTF psychological treatment designed specifically for postnatal women with diagnosed depressive disorders.

Aims of This Research

This study aims to evaluate remission from clinical depression and reduction of depressive and anxiety symptom severity among a sample of postnatal women with diagnosed depression. We aim to evaluate the relative efficacy for women randomly assigned to (1) MMB, a guided version of an internet intervention for PND; (2) a validated FTF-CBT program for PND; and (3) TAU. The secondary aim of this randomized controlled trial (RCT) is to compare the 2 active conditions with respect to symptom trajectory, stress, and process measures of treatment engagement, competence in the mothering role, marital functioning, acceptability, and satisfaction.

Our a priori hypothesis was that women allocated to either of the 2 active conditions (MMB and FTF-CBT) would have significantly higher remission rates, greater reductions in depression and anxiety symptoms, and greater improvement in secondary outcomes compared with the TAU condition but that MMB and FTF-CBT would not differ significantly from each other.

Methods

Experimental Design

This study was a parallel 3-group RCT involving 116 participants and consistent with the CONSORT standards [53,54]. The trial was prospectively registered in the Australia and New Zealand Clinical Trials Registry (trial registration number ACTRN12613000881730) and was approved by the human research ethics committee of Austin Health, Melbourne (approval number H2013/04972).

Procedure

Screening and recruitment were conducted via maternal and child health centers in rural and metropolitan Victoria, Australia, and by localized advertising on the internet, in newspapers, and on the radio. The project was conducted between August 2014 and November 2017.

Inclusion criteria were determined in 2 phases. Phase 1 criteria included the following: Edinburgh Postnatal Depression Scale [55] scores of 11-25, aged ≥18 years, 6 weeks to 1 year postpartum; home internet access, familiarity with internet and email, and able and willing to give informed consent that included agreeing to be assigned to any of the 3 experimental conditions. Women who scored >0 on item number 10 of the Edinburgh Postnatal Depression Scale (thoughts of self-harm) were asked follow-up questions to ascertain intentionality, plan, lethality, access to means, and history of suicide attempts. Women deemed to be at risk of suicide were excluded and referred to receive immediate crisis attention.

Women found eligible in the screening phase 1 were subsequently assessed in phase 2 by a clinical psychologist or a provisional psychologist in a phone-administered SCID-IV [56]. Prospective study participants were excluded if they met any of the following exclusion criteria in screening phase 2: current substance abuse, manic or hypomanic symptoms or depression with psychotic features meeting the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*) criteria, posttraumatic stress disorder, risk of suicide, and under current treatment for depression (medication or psychotherapy). Women who received a diagnosis of major or minor depressive episodes and who met all other eligibility criteria were invited to participate. Women who gave written consent and completed the baseline assessment were then randomized to a study condition.

Randomization

Women were randomized in a 1:1:1 ratio to 1 of the 3 conditions: internet-based CBT with telephone support (MMB; 39/116, 33.6%), individual FTF-CBT (39/116, 33.6%), or TAU (38/116, 32.8%). An automated permuted block (block sizes of 3, 6, and 9) allocation schedule was pregenerated by a computer by an independent technician. Allocation concealment was ensured by a central, computer-automated administration. Upon completion of baseline questionnaires, the system assigned the participants to 1 of the 3 conditions. Their allocation was subsequently revealed to them in a phone call. Given the nature of the treatment conditions, allocation could not be concealed from participants beyond the point of allocation.

Experimental Conditions

MumMoodBooster

MMB is designed to deliver content that is similar to FTF depression treatment, with tailored, interactive activities used to address individual issues and engage women. Support from a telephone coach is intended to encourage women to use and complete the program. The initial steps of the program provide explicit direction, whereas the latter steps encourage participants

to assume increasingly greater responsibility for managing their own plan for change.

MMB uses tunnel information architecture [41,57] with step-by-step guidance through 6 sessions, with a new session becoming available for use every week. Session content is similar to that found in traditional FTF-CBT treatment using a scaffold approach that builds upon concepts presented in the previous sessions. MMB's charting function was designed to help participants see the functional relationship between mood and activity levels—a key therapeutic concept in CBT that can be maximized in an internet program and, therefore, an important functionality in this program. Information from past sessions is used to reinforce gains made, and program content is tailored to individual issues with the provision of ipsative feedback [58]. A printable summary describes the key content covered in each session and a tailored list of recommended home practice activities. MMB includes content designed to enhance participant self-efficacy to accomplish recommended strategies as well as text, audios, and videos to engage participants in a program with some personalized elements. In recognition of the important role of partners and paternal depression in the treatment process of PND [59], an article in the MMB library entitled *You and Your Partner* provides content relevant to partners of women with PND. In addition, MMB enables participants to choose whether to invite their partner to use a free-standing partner support website with a separate user log-in. The MMB library also includes articles on relaxation, problem solving, and getting support for parenting (for further details on the development of the program, see the study by Danaher et al [41]). As described in earlier publications [42,43,60,61], to encourage optimal engagement and resulting behavior change [57], although the intervention can be accessed and used on smartphones, MMB was designed and optimized for use on desktop computers, laptops, and tablets and designed to function on popular browsers for both Windows PCs and Mac computers.

Weekly low-intensity telephone coaching support (30 minutes maximum per week) was provided. Rather than providing therapy per se, coaches were instructed to reinforce participant progress, encourage program use (practice of strategies and completion of tasks), and introduce the themes of upcoming sessions. In an initial welcome call, the assigned MMB coach explained the MMB core structure, additional library articles, and partner website. Coaches were able to access a secure administrative website to review each participant's program use to help tailor their support. Coaching call fidelity was facilitated by the use of a manualized script and a session-by-session checklist. Coaching calls were provided over a period of 9 weeks, which enabled participants to work through the 6 sessions at <1 session per week and allowed for the rescheduling of up to 3 missed coach calls.

The FTF-CBT Condition

Women in the FTF-CBT condition were scheduled to receive weekly individualized CBT therapy from an experienced psychologist who followed a detailed, scripted manual for delivery of the *Getting Ahead of Postnatal Depression* program [13,62]. This is a manualized 9-session CBT-based program whose core content has been shown to be effective in several RCTs [13,63,64], and this was drawn upon in the original design of MMB. The program addressed maternal mood (depression and anxiety), behavioral activation, cognitive strategies, self-esteem, relaxation (*relax on the run*), getting support, and dealing with partner issues. The program included 1 additional session that involved both participants and their partners [62].

After screening was completed on the web, the research protocol was explained to potential recruits and following a SCID-IV interview, if randomized to FTF-CBT, participants visited their general practitioner (GP) with a referral letter from the study to arrange treatment with a suitable local psychology provider located by the research team. Psychologists kept session-by-session checklists of items covered in the FTF-CBT program. A report was also sent to the GPs of participants in the other 2 conditions; however, their GPs were not actively encouraged to connect their patients with a local psychology provider.

The TAU Condition

Women in the TAU condition were referred back to their GP supplemented with a written summary of their diagnostic assessment. Support and referral to other services could then occur as necessary as typically occurs in Australia when specialized programs are not available.

Measures

Measures were collected primarily using web-based questionnaires, although phone calls were also used, as described in the following section (Table 1). All participants received automated email prompts reminding them to complete the web-based assessments. Participants were given a modest reimbursement in consideration of their time spent in completing questionnaires and assessments at the 12-week posttest and 21-week follow-up: Aus $20 (US $14.90) and Aus $35 (US $26), respectively.

Table 1. Measures and collection time points.

Measure	Baseline enrollment	Safety monitoring			Posttest, week 12	Week 16 (safety monitoring)	Follow-up, week 21
		Week 3	Week 5	Week 9			
Depressive symptoms							
Structured Clinical Interview for DSM-IV[a]	✓[b]						✓
Beck Depression Inventory	✓				✓		✓
Patient Health Questionnaire	✓	✓	✓	✓	✓	✓	✓
Anxiety symptoms							
Depression Anxiety Stress Scale	✓			✓	✓		✓
Perceived stress							
Depression Anxiety Stress Scale	✓			✓	✓		✓
Marital functioning							
Dyadic Adjustment Scale	✓				✓		✓
CBT[c] skills							
Negative thinking: Automatic Thoughts Questionnaire	✓				✓		✓
Behavioral activation: Behavioral Activation for Depression Scale	✓				✓		✓
Maternal self-efficacy							
Parenting Sense of Competence Scale Maternal Self-efficacy subscale	✓				✓		✓
Monitoring and safety measures							
Risk management protocol	✓	✓	✓	✓	✓	✓	✓
Treatment satisfaction							✓
Use of other supports or treatments							✓

[a]DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

[b]Measurement collected.

[c]CBT: cognitive behavioral therapy.

Primary Outcome Measures

Trained diagnostic interviewers blinded to treatment allocation conducted the SCID-IV [56] by phone to determine a *DSM-IV* diagnosis of major depression or minor depression [65]. To measure the severity of depression, we used a web-based version of the revised Beck Depression Inventory (BDI-II) [66,67], a well-validated, 21-item clinical instrument that measures cognitive, affective, and physiological factors. The BDI-II has been used in numerous studies involving perinatal women [68-70] and has been validated against diagnostic criteria in perinatal populations [69]. The 9-item Patient Health Questionnaire (PHQ-9) [71] was used as a serial measure of depression for trajectory analysis. The PHQ-9 measures depressive symptoms in the past 2 weeks; it (1) is well-validated [71,72] (including with perinatal women [73]), (2) has high test–retest reliability [74], (3) has high internal consistency (α=.86) [75], and (4) is shown to be sensitive to changes in response to treatment [76]. Participants' anxiety symptom severity was measured using the 7-item Anxiety Scale of the Depression Anxiety Stress Scale, 21-item (DASS-21) [77,78]. The DASS-21 uses a 4-point scale to describe agreement with statements over the past week and is rated as 0=did not apply to me at all, never; 1=applied to me to some degree or some of the time, sometimes; 2=applied to me to a considerable degree or a good part of the time, often; and 3=applied to me very much or most of the time, almost always. The sum of responses to the Anxiety Scale is scored as normal (0-7), mild (8-9), moderate (10-14), severe (15-19), and extremely severe (>20). The DASS-21 has good concurrent validity with other established anxiety scales [77].

Secondary Outcome Measures

CBT Skills: Negative Thinking

Participants completed the 30-item Automatic Thoughts Questionnaire (ATQ) [79]. The ATQ asks respondents to rate their agreement with a series of statements (eg, "My life is a mess") using a 5-point scale (1=not at all to 5=all of the time). The maximum ATQ score is 150.

CBT Skills: Behavioral Activation

Participants completed the 25-item Behavioral Activation for Depression Scale [80], which measures changes in activation, avoidance or rumination, work or school impairment, and social impairment. Respondents rated their agreement with a series of statements (eg, "I stayed in bed for too long even though I had things to do") on a scale of 0-6. The maximum Behavioral Activation for Depression Scale score is 150.

Maternal Self-efficacy

Drawing upon Bandura's theoretical work, Wittkowski et al [81] described parenting self-efficacy as a parent's belief in their ability to perform the parenting role successfully. To measure this, we used the Maternal Self-efficacy subscale of the Parenting Sense of Competence Scale [82], which asks respondents to rate the extent of their agreement with 7 items relating to self-perception of knowledge and competence in the mothering role. Each statement (eg, "I honestly believe I have all the skills necessary to be a good mother to my baby") is rated from 1 (strongly disagree) to 6 (strongly agree). The maximum score is 42.

Marital Functioning

Women's relationships with their partners were assessed using the 7-item Dyadic Adjustment Scale [83]. The general satisfaction score was calculated as the sum of all items. The maximum score is 36.

Treatment Satisfaction and Helpfulness

Participants in the MMB and FTF-CBT conditions were asked to rate their satisfaction with their assigned treatment as well as a set of select treatment components using a 4-point scale (0=not at all satisfied, 1=somewhat satisfied, 2=moderately satisfied, and 3=very satisfied). MMB participants were also asked to rate the helpfulness of phone coach calls using a 4-point scale (0=not at all helpful, 1=somewhat helpful, 2=moderately helpful, and 3=very helpful).

Use of Other Supports or Treatments

Participants were asked to complete a checklist of other support services and treatments they used during the study interval.

Participant Engagement

MMB use (overall number and duration of discrete visits from enrollment to the end of the follow-up assessment) was tracked continuously and unobtrusively via database flags and analysis of server activity logs (see our earlier MMB trial publications for more details [41-43]).

Coaches in the MMB condition and psychologists in the FTF-CBT condition recorded data on participant session completion on session-by-session compliance checklists itemizing the elements of each session plan covered. A subset of coaching calls and FTF-CBT therapy sessions were audio recorded (with permission) to facilitate fidelity checks.

All measures were collected at baseline, 12-week posttest, and 21-week follow-up, except for the SCID-IV diagnostic assessment, which was not conducted at 12 weeks. In addition, the PHQ-9 was also monitored during safety calls at weeks 3, 5, 9, and 16 after enrollment. The follow-up time point of 21 weeks was selected to reflect 3 months of posttreatment completion (with MMB and FTF-CBT treatment taking 9 weeks to complete).

Safety Monitoring

Symptom severity and safety of participants in each of the RCT conditions were monitored in calls for all conditions using the PHQ-9. A written risk management protocol, based on the successful approach of Simon et al [84], was initiated if any risk of harm to self or infant or marked deterioration of depressive symptoms was indicated. These assessments of symptom severity and safety were scripted and accomplished on the phone by trained research staff who were blind to the respondents' treatment allocation. In the 2 intervention conditions, weekly contact with therapists and coaches enabled additional monitoring of depression severity and safety.

Data Analysis

Depression remission was assessed using the evidence-based definition of Frank et al [85], which requires a person to be asymptomatic (defined as no longer meeting the criteria for the disorder and having ≤2 symptoms for ≥2 weeks).

Categorical outcomes (eg, diagnostic status based on the SCID-IV) were analyzed using contingency tables and logistic regression; survival analysis (Cox regression and Kaplan–Meier product limit) was used to predict time-to-remission of the index depressive episode using the results from follow-up interviews. Continuous outcomes (eg, symptom severity) were analyzed using random-effects regression models, accommodating time-independent and time-dependent covariates, fixed and random factors, and incomplete data.

Random-effects growth models were estimated from baseline (the intercept) to the 21-week follow-up using the PROC MIXED procedure in SAS 9.2 (SAS Institute). The growth models were estimated with full information maximum likelihood and an unstructured variance/covariance matrix, and the intercepts and slopes were allowed to vary. First, unconditional means models were run for each outcome and the estimated and observed data plotted. Next, linear conditional growth models were run with time coded in weeks since the pretest assessment. All primary analyses to address major hypotheses adhered to intention-to-treat principles, and 3 a priori comparisons were examined: MMB versus TAU, FTF-CBT versus TAU, and MMB versus FTF-CBT (for each comparison, the first group was the reference group). The condition×time interaction is a test of the efficacy of the program for each a priori comparison. The effect size for the condition×time interaction was computed as a d-statistic equivalent [86].

Descriptive statistics and plots were used to screen all outcomes for normality and outliers. No serious violation of normality was found, with the exception of anxiety at 12 and 21 weeks, which showed a preponderance of zeros. During modeling, anxiety scores were examined in their normal metric and also log transformed to better approximate normality. A participant was flagged as an outlier on anxiety and stress at the 12-week assessment. During modeling, the outlier was evaluated for its impact on change over time.

Clinical Significance, Power, and Sample Size

An a priori consideration of power suggested that for the survival analysis predicting time-to-remission from the index episode, with a 2-tailed α of .05, n=70 per condition would yield 80% power to detect a small effect size (hazard ratio 2.0). For the measure of depressive symptoms (BDI-II), data from previous work by Milgrom et al [13] provided a baseline value=23, σ=8.09. A difference of 6.5 takes scores below the threshold of *clinical* depression (ie, a score <17 points). With 80% power at α=.05, the required $n=15.7\times(8.09/6.5)^2=24.3$, which rounds up to 25. Taking into account a prudent noncompliance rate of 30%, the adjusted sample size, $n^* = 25 / (1 - 0.3)^2 = 51$, which rounds up to 55. Thus, an estimated n=70 per condition yielded sufficient power to detect a small between-group effect size and a minimum clinically important difference in the primary measure of symptom severity. This sample size was not achieved within the funding period; however, post hoc analysis of power with the sample size achieved confirmed sufficient power to address the minimum clinically important difference for the primary outcomes. We had sufficient power (>80%) to detect significant group differences ($P<.05$) in our primary outcomes. We used a simulation approach with 1000 replications using the likelihood ratio test [87] and implemented it in SAS. We estimated the power to detect condition effects for changes in the primary outcomes. Unlike the a priori power estimates, this post hoc power estimation included observed parameter estimates (eg, sample size, effect size, and correlations between intercept and slope). Thus, this study was sufficiently powered to detect group differences between MMB and TAU for medium to large effects.

Depressive Symptom Trajectory

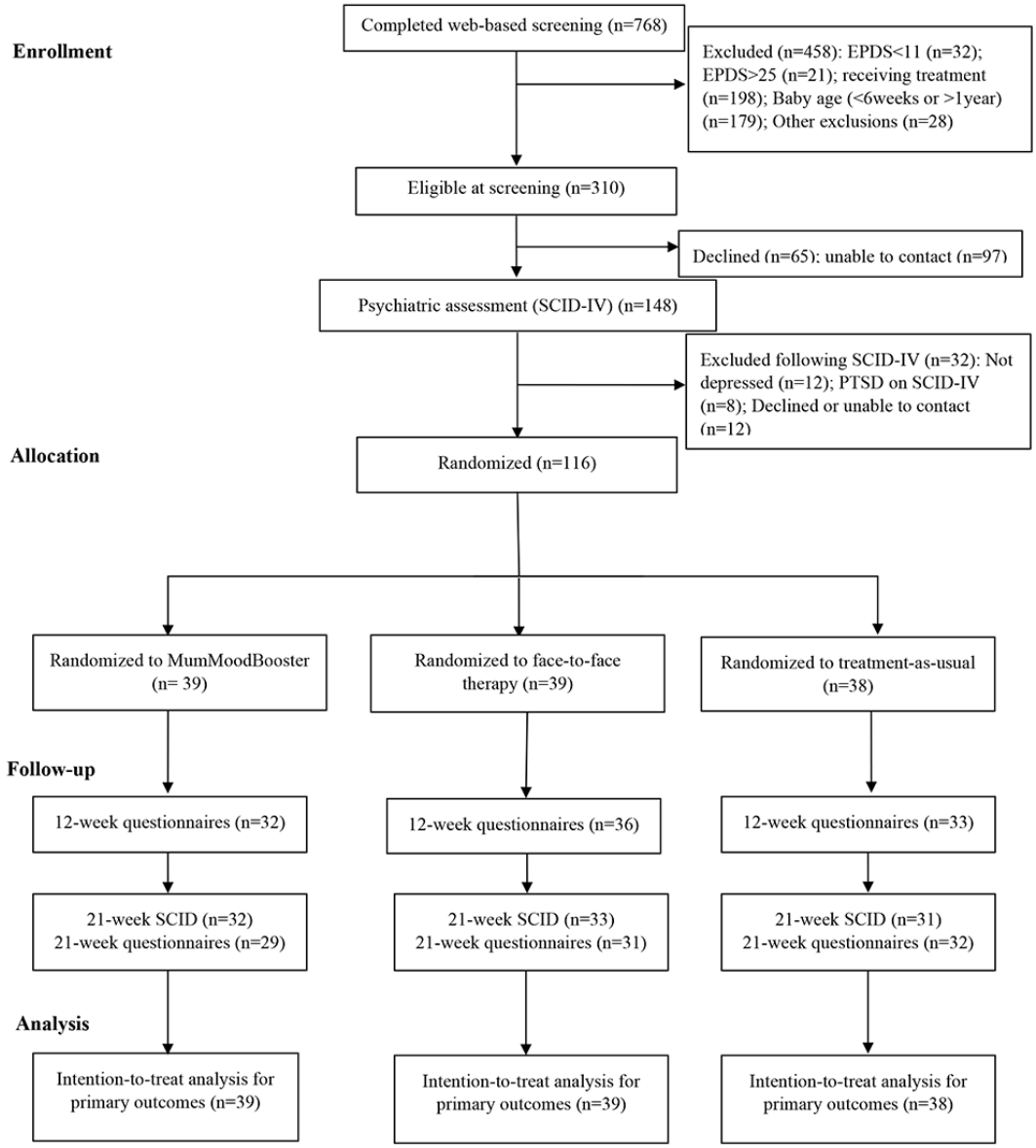
At baseline and during the routinely scheduled safety monitoring calls (at weeks 3, 5, 9, 12, and 21), all participants completed the 9-item PHQ-9 over the telephone. These data were used to calculate the trajectory of depressive symptom changes over time.

Results

Sample

A total of 116 participants were randomized during the study period. Figure 1 presents a CONSORT chart of participant flow through the trial.

Figure 1. Participant flow through the study. EPDS: Edinburgh Postnatal Depression Scale; PTSD: posttraumatic stress disorder; SCID-IV: Structured Clinical Interview for the DSM-IV.



Participant Characteristics

A range of primiparous and multiparous, high-income and low-income women were included (see participant baseline characteristics in Table 2). On average, participants were aged 32.1 years (SD 4.7); 66.4% (77/116) of participants were married, and regarding the highest level of education, 12.2% (14/115) completed high school, 26.1% (30/115) had completed a bachelor's degree, and 34.8% (40/115) had completed graduate or postgraduate programs. The average infant age was 26.1 (SD 14.5) weeks. Chi-square tests and one-way analysis of variance were used to examine baseline demographic characteristics by condition. No group differences were found, confirming that randomization produced initially equivalent groups.

Table 2. Participant and baby characteristics by study condition (N=116)[a].

Participant characteristics	MMB[b] (n=39)	FTF-CBT[c] (n=39)	TAU[d] (n=38)
Age (years), mean (SD)	30.8 (4.3)	32.2 (5.3)	31.9 (4.2)
Number of children, n (%)			
1	20 (51)	19 (49)	16 (42)
2	17 (44)	15 (38)	16 (42)
3	2 (5)	3 (8)	5 (13)
≥4	0 (0)	2 (5)	1 (3)
Marital status, n (%)			
Married	27 (69)	26 (67)	24 (63)
De facto	12 (31)	10 (26)	11 (29)
Separated	0 (0)	1 (3)	2 (5)
Single	0 (0)	2 (5)	1 (3)
Highest education, n (%)			
Completed high school			
No	1 (3)	1 (3)	1 (3)
Yes	4 (10)	6 (16)	4 (11)
Certificate or apprenticeship	7 (18)	7 (18)	8 (21)
Advanced diploma	4 (10)	1 (3)	1 (3)
Bachelors	7 (18)	12 (32)	11 (29)
Graduate	7 (18)	5 (13)	4 (11)
Postgraduate	9 (23)	6 (16)	9 (24)
Annual household income Aus $ (US $), n (%)			
>20,000 (14,800)	0 (0)	1 (3)	0 (0)
20,001-40,000 (14,801-29,600)	0 (0)	6 (15)	3 (8)
40,001-60,000 (29,601-44,400)	4 (10)	3 (8)	5 (13)
60,001-80,000 (44,401-59,200)	8 (21)	5 (13)	7 (18)
≥80,001 (59,201)	24 (62)	21 (54)	22 (58)
Prefer not to disclose	3 (8)	3 (8)	1 (3)
Prior counseling, n (%)	26 (67)	26 (67)	20 (53)
Taken depression medication, n (%)	9 (23)	15 (38)	7 (18)
Baby characteristics			
Age (weeks), mean (SD)	28.5 (14.1)	31.3 (16.1)	28.4 (13.7)
Female, n (%)	25 (64)	24 (62)	18 (47)

[a]Column totals in percentages may not sum to 100; values in percentages are rounded upwards to the nearest full integer.

[b]MMB: MumMoodBooster.

[c]FTF-CBT: face-to-face cognitive behavioral therapy.

[d]TAU: treatment as usual.

Missing Data and Attrition

Failure to complete scheduled assessments (attrition from the study) was 12.9% (15/116) at the 12-week assessment and 20.7% (24/116) at the 21-week assessment (Figure 1). Analysis indicated that attrition at these times was not related to study condition (P=.57), any of the demographic characteristics in Table 1 (all P>.22), or baseline measures of the primary or secondary outcomes (all P>.07).

Engagement in MMB Program

Of the women allocated to the MMB intervention, 85% (33/39) completed ≥3 MMB sessions, and 72% (28/39) viewed all 6 MMB sessions. The MMB program was not visited by 3% (1/39) of participants. Of the women who visited MMB at least

once, the mean number of sessions viewed was 5.6 (SD 1.7), and the mean number of visits was 15.6 (SD 8.7). The mean total time spent using MMB averaged 230 (range 2-440) minutes. The mean number of library articles accessed was 3.5 out of a possible 8, and the mean number of activities accessed was 15 out of 63. Approximately 39% (15/39) of participant partners accessed the partner support website.

Engagement in FTF-CBT

Of the women offered FTF-CBT, 62% (24/39) completed ≥3 sessions, 51% (20/39) finished ≥6 sessions, and 46% (18/39) completed all 10 sessions. Approximately 31% (12/39) of women did not attend any FTF sessions. The mean number of FTF-CBT sessions attended for all 39 women was 5.5 (SD 4.5).

Primary Outcomes

At baseline, all women in all 3 groups had a *DSM-IV* diagnosis of either major or minor depression; 92.2% (107/116) of the sample had a major depression diagnosis. Table 3 shows the change in the profile of depressive diagnostic status from baseline to the 21-week follow-up. On the basis of the observed data available at the 21-week follow-up assessment, the FTF-CBT group was 56.6% less likely to have a diagnosis of major or minor depression relative to the TAU group. The MMB group was 47.7% less likely to have a diagnosis of major or minor depression relative to the TAU group. Remission rates in the FTF-CBT and MMB conditions were not significantly different from each other (χ^2_1=0.1, *P*=.81). Table 4 provides descriptive statistics for the primary continuous outcome measures for each of the 3 conditions at baseline, 12 weeks, and 21 weeks, and Table 5 provides descriptive statistics for the secondary outcome measures.

Table 3. Diagnostic status of participants from baseline to the 21-week follow-up (N=116).

Conditions	Baseline, n (%)		21-week follow-up, n (%)			
	Major	Minor	Major	Minor	Lost to follow-up	Remission[a]
MMB[b] (n=39)	36 (92)	3 (8)	4 (10)	3 (8)	7 (18)	25 (78)
FTF-CBT[c] (n=39)	35 (90)	4 (10)	5 (13)	1 (3)	6 (15)	27 (82)
TAU[d] (n=38)	36 (95)	2 (5)	12 (32)	1 (3)	7 (18)	18 (58)

[a]Remission was defined as the absence of a Structured Clinical Interview for the DSM-IV diagnosis for major or minor depression at the follow-up interview and excludes cases lost to follow-up.

[b]MMB: MumMoodBooster.

[c]FTF-CBT: face-to-face cognitive behavioral therapy.

[d]TAU: treatment as usual.

Table 4. Descriptive statistics for continuous primary outcomes.

Primary outcome	Baseline, mean (SD)	12-week, mean (SD)	21-week, mean (SD)
Depressive symptoms (BDI-II)[a,b]			
MMB[c]	28.10 (7.91)	11.63 (8.96)	8.70 (6.92)
FTF-CBT[d]	27.18 (9.95)	21.36 (12.15)	15.00 (10.71)
TAU[e]	29.97 (8.76)	18.85 (10.16)	17.41 ()11.51
Anxiety symptoms (DASS-21)[f,g]			
MMB	8.87 (7.02)	1.81 (2.61)	2.69 (4.29)
FTF-CBT	9.23 (7.31)	8.11 (8.01)	4.45 (5.45)
TAU	8.58 (7.85)	4.73 (5.22)	5.19 (5.52)

[a]BDI-II: Beck Depression Inventory, revised.

[b]BDI-II categories: minimal depression=0-13, mild depression=14-19, moderate depression=20-28, and severe depression=29-63.

[c]MMB: MumMoodBooster.

[d]FTF-CBT: face-to-face cognitive behavioral therapy.

[e]TAU: treatment as usual.

[f]DASS-21: Depression Anxiety Stress Scale, 21-item.

[g]DASS-21 categories: normal=0-7, mild=8-9, moderate=10-14, severe=15-19, and extremely severe=≥20.

Table 5. Secondary outcomes.

Secondary outcome	Baseline, mean (SD)	12-week, mean (SD)	21-week, mean (SD)
Perceived stress (DASS-21)[a,b]			
MMB[c]	20.67 (7.99)	10.88 (7.22)	8.55 (6.88)
FTF-CBT[d]	18.72 (8.10)	16.94 (9.15)	12.39 (7.97)
TAU[e]	20.16 (7.52)	14.67 (6.28)	13.25 (8.11)
Marital functioning (DAS-7[f])			
MMB	21.23 (5.22)	21.25 (4.56)	23.21 (6.56)
FTF-CBT	19.90 (7.66)	21.39 (6.18)	21.07 (7.87)
TAU	21.79 (6.66)	22.49 (6.13)	22.22 (6.41)
CBT[g] skills-behavioral activation (BADS[h])			
MMB	55.69 (18.12)	45.06 (17.52)	42.93 (14.13)
FTF-CBT	57.03 (20.14)	50.78 (15.19)	40.84 (11.13)
TAU	58.44 (15.92)	52.36 (16.15)	48.53 (16.53)
CBT skills-negative thinking (ATQ[i])			
MMB	47.72 (25.69)	21.44 (21.40)	16.52 (18.02)
FTF-CBT	45.77 (28.65)	36.25 (27.74)	24.55 (23.60)
TAU	52.84 (25.70)	34.15 (24.62)	32.66 (27.11)
Maternal Self-efficacy subscale (PSOC[j])			
MMB	24.03 (6.58)	29.66 (7.39)	30.90 (7.87)
FTF-CBT	25.61 (7.96)	26.44 (7.12)	29.71 (7.26)
TAU	22.45 (7.21)	28.03 (7.20)	28.09 (7.83)

[a]DASS-21: Depression Anxiety Stress Scale, 21-item.

[b]DASS-21 categories: normal=0-14, mild=15-18, moderate=19-25, severe=26-33, and extremely severe=≥34.

[c]MMB: MumMoodBooster.

[d]FTF-CBT: face-to-face cognitive behavioral therapy.

[e]TAU: treatment as usual.

[f]DAS-7: Dyadic Adjustment Scale.

[g]CBT: cognitive behavioral therapy.

[h]BADS: Behavioral Activation for Depression Scale.

[i]ATQ: Automatic Thoughts Questionnaire.

[j]PSOC: Parenting Sense of Competence Scale.

Growth Models

Parameter estimates from the growth models for each a priori comparison are provided in Tables 6-8. These include parameter estimates for intercept, condition, time, and condition×time interaction.

In Table 6, the intercept reflects the model-implied estimated baseline depressive score (PHQ-9) for the TAU group (29.02). The *t* value and *P* value indicate that the intercept is significantly different from 0 (*P*<.001). The condition parameter (−2.154) indicates that the MMB condition had an estimated baseline score that was 2.15 points lower than the TAU group. The *t* value and *P* value indicate that the lower baseline depressive score for the MMB condition was not statistically significant (*P*=.24). The time parameter (−0.614) indicates that the TAU group had an estimated linear decrease in depressive symptom score of 0.61 points for each week through the 21-week follow-up assessment. The *t* value and *P* value indicate that the weekly linear decrease for the TAU group is significantly different from 0 (*P*<.001). The condition×time interaction term (−0.3029) indicates that the MMB group decreased an additional 0.30 points each week through the 21-week follow-up. The *t* value and *P* value indicate that the greater decrease over time for the MMB group relative to the TAU group was statistically significant (*P*=.01). Figure 2 shows the trajectory of the PHQ-9 scores for each condition over the course of the study. Significant condition×time effects were also found for anxiety and perceived stress (Table 6), indicating that MMB was statistically superior to TAU in reducing symptoms of depression, anxiety, and stress from baseline to the 21-week follow-up.

Table 6. Parameter estimates from growth models for MumMoodBooster (MMB) versus treatment as usual (TAU).

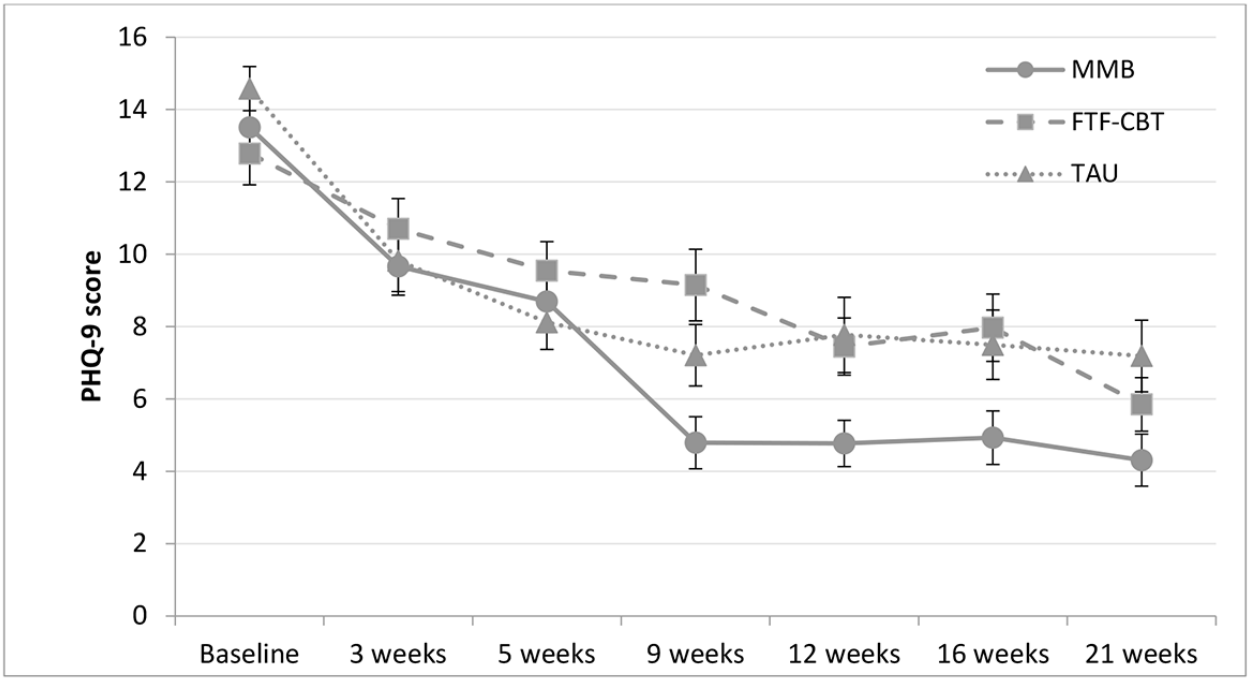
Outcomes and growth model parameters	Parameter (SE)	*t* test[a] (*df*)	*P* value
Depressive symptoms			
Intercept	29.024 (1.301)	22.31 (75)	<.001
Condition	−2.154 (1.831)	−1.18 (75)	.24
Time	−0.614 (0.088)	−7.01 (124)	<.001
Condition×time	−0.329 (0.126)	−2.61 (124)	.01
Anxiety symptoms			
Intercept	8.144 (1.129)	7.21 (75)	<.001
Condition	0.012 (1.588)	0.01 (75)	.99
Time	−0.163 (0.052)	−3.15 (124)	.002
Condition×time	−0.148 (0.074)	−2.01 (124)	.05
Perceived stress			
Intercept	19.799 (1.159)	17.08 (75)	<.001
Condition	0.186 (1.631)	0.11 (75)	.91
Time	−0.323 (0.068)	−4.78 (124)	<.001
Condition×time	−0.270 (0.097)	−2.78 (124)	.006
Marital functioning			
Intercept	21.906 (0.930)	23.55 (75)	<.001
Condition	−0.898 (1.308)	−0.69 (75)	.50
Time	0.030 (0.042)	0.72 (124)	.47
Condition×time	0.068 (0.060)	1.13 (124)	.26
CBT[b] skills: behavioral activation			
Intercept	58.158 (2.651)	21.94 (75)	<.001
Condition	−3.258 (3.728)	−0.87 (75)	.39
Time	−0.466 (0.137)	−3.40 (124)	.001
Condition×time	−0.134 (0.197)	−0.68 (124)	.50
CBT skills: negative thinking			
Intercept	51.106 (3.992)	12.80 (75)	<.001
Condition	−5.410 (5.614)	−0.96 (75)	.34
Time	−0.996 (0.185)	−5.40 (124)	<.001
Condition×time	−0.408 (0.265)	−1.54 (124)	.13
Maternal Self-efficacy subscale (PSOC[c])			
Intercept	23.021 (1.092)	21.08 (75)	<.001
Condition	1.364 (1.535)	0.89 (75)	.38
Time	0.274 (0.059)	4.65 (124)	<.001
Condition×time	0.077 (0.085)	0.92 (124)	.36

[a]Tests were two-tailed.

[b]CBT: cognitive behavioral therapy.

[c]PSOC: Parenting Sense of Competence Scale.

Figure 2. Change from baseline to 21-week follow-up for the MMB, FTF-CBT and TAU groups on the PHQ-9. Mean values are observed values plotted ±1 SE. PHQ-9 score ranges: minimal or none=0-4, mild=5-9, moderate=10-14, moderately severe=15-19, and severe ≥20. FTF-CBT: face-to-face cognitive behavioral therapy; MMB: MumMoodBooster; PHQ-9: Patient Health Questionnaire–9; TAU: treatment as usual.



Tables 7 and 8 show the results for the other a priori comparisons: FTF-CBT versus TAU and MMB versus FTF-CBT. Although none of the condition×time effects were significant for the FTF-CBT versus TAU comparison (Table 7), the MMB versus FTF-CBT comparison revealed that MMB performed significantly better than FTF-CBT in reducing depressive symptoms and perceived stress (Table 8).

The effect size (*d*) for all the condition×time interaction terms is presented in Table 9. Compared with TAU, the MMB group reported a large reduction in depressive symptoms (*d*=−0.83) and medium-sized reductions in anxiety (*d*=−0.42) and perceived stress (*d*=−0.73). Compared with the FTF-CBT condition, MMB participants reported large reductions in both depressive symptoms (*d*=−0.98) and perceived stress (*d*=−0.89). The TAU and FTF-CBT conditions did not differ significantly in any of the outcome measures.

Table 7. Parameter estimates from growth models of face-to-face cognitive behavioral therapy (CBT) versus treatment as usual.

Outcomes and growth model parameters	Parameter (SE)	*t* test[a] (*df*)	*P* value
Depressive symptoms			
Intercept	29.047 (1.533)	18.94 (75)	<.001
Condition	−1.733 (2.155)	−0.80 (75)	.42
Time	−0.611 (0.094)	−6.52 (124)	<.001
Condition×time	0.110 (0.132)	0.83 (124)	.41
Anxiety symptoms			
Intercept	8.140 (1.215)	6.70 (75)	<.001
Condition	1.389 (1.707)	0.81 (75)	.42
Time	−0.155 (0.054)	−2.90 (124)	.004
Condition×time	−0.012 (0.076)	−0.16 (124)	.88
Perceived stress			
Intercept	19.799 (1.218)	16.26 (75)	<.001
Condition	−0.705 (1.711)	−0.41 (75)	.68
Time	−0.314 (0.060)	−5.23 (124)	<.001
Condition×time	0.068 (0.085)	0.80 (124)	.42
Marital functioning			
Intercept	21.911 (1.138)	19.26 (75)	<.001
Condition	−1.874 (1.599)	−1.17 (75)	.25
Time	0.029 (0.050)	0.57 (124)	.57
Condition×time	−0.012 (0.070)	−0.17 (124)	.87
CBT skills: behavioral activation			
Intercept	−0.420 (4.015)	−0.10 (75)	.92
Condition	−0.420 (4.015)	−0.10 (75)	.92
Time	−0.459 (0.142)	−3.24 (124)	.002
Condition×time	−0.217 (0.200)	−1.09 (124)	.28
CBT skills: negative thinking			
Intercept	51.091 (4.272)	11.96 (75)	<.001
Condition	−4.741 (6.002)	−0.79 (75)	.43
Time	−0.989 (0.204)	−4.84 (124)	<.001
Condition×time	0.075 (0.288)	0.26 (124)	.80
Maternal Self-efficacy subscale (PSOC[c])			
Intercept	23.015 (1.204)	19.11 (75)	<.001
Condition	2.366 (1.692)	1.40 (75)	.17
Time	0.278 (0.066)	4.19 (124)	<.001
Condition×time	−0.096 (0.093)	−1.03 (124)	.31

[a]Tests were two-tailed.

[b]PSOC: Parenting Sense of Competence Scale.

Table 8. Parameter estimates from growth models for MumMoodBooster versus face-to-face cognitive behavioral therapy (CBT).

Outcomes and growth model parameters	Parameter (SE)	*t* test[a] (*df*)	*P* value
Depressive symptoms			
Intercept	27.373 (1.446)	18.93 (76)	<.001
Condition	−0.491 (2.047)	−0.24 (76)	.81
Time	−0.525 (0.086)	−6.12 (126)	<.001
Condition×time	−0.418 (0.123)	−3.39 (126)	.001
Anxiety symptoms			
Intercept	9.567 (1.108)	8.64 (76)	<.001
Condition	−1.419 (1.568)	−0.90 (76)	.37
Time	−0.182 (0.057)	−3.17 (126)	.002
Condition×time	−0.129 (0.082)	−1.57 (126)	.12
Perceived stress			
Intercept	19.109 (1.236)	15.46 (76)	<.001
Condition	0.879 (1.750)	0.50 (76)	.62
Time	−0.252 (0.064)	−3.96 (126)	<.001
Condition×time	−0.341 (0.092)	−3.73 (126)	<.001
Marital functioning			
Intercept	20.074 (0.977)	20.55 (76)	<.001
Condition	0.932 (1.383)	0.67 (76)	.50
Time	0.004 (0.044)	0.08 (126)	.94
Condition×time	0.097 (0.064)	1.52 (126)	.13
CBT skills: behavioral activation			
Intercept	57.725 (2.917)	19.79 (76)	<.001
Condition	−2.800 (4.130)	−0.68 (76)	.50
Time	−0.695 (0.145)	−4.78 (126)	<.001
Condition×time	0.082 (0.207)	0.40 (126)	.69
CBT skills: negative thinking			
Intercept	46.395 (4.291)	10.81 (76)	<.001
Condition	−0.640 (6.073)	−0.11 (76)	.92
Time	−0.935 (0.204)	−4.58 (126)	<.001
Condition×time	−0.499 (0.293)	−1.71 (126)	.09
Maternal self-efficacy subscale (PSOC[b])			
Intercept	25.375 (1.152)	22.03 (76)	<.001
Condition	−0.987 (1.630)	−0.61 (76)	.55
Time	0.188 (0.058)	3.24 (126)	.002
Condition×time	0.162 (0.084)	1.94 (126)	.06

[a]Tests were two-tailed.

[b]PSOC: Parenting Sense of Competence Scale.

Table 9. Effect sizes (d) and significance levels for condition×time differences from mixed-effects growth models[a].

Outcome measures	MMB[b] vs TAU[c]		FTF-CBT[d] vs TAU		MMB vs FTF-CBT	
	d	*P* value	*d*	*P* value	*d*	*P* value
Depressive symptoms	–0.83	.01[e]	0.25	.41	–0.98	.001[e]
Anxiety symptoms	–0.42	.05[e]	–0.03	.88	–0.38	.12
Perceived stress	–0.73	.006[e]	0.18	.42	–0.89	<.001[e]
Marital functioning	0.2	.26	–0.03	.87	0.29	.13
CBT[f] skills						
Behavioral activation	–0.16	.50	–0.25	.28	0.09	.69
Negative thinking	–0.33	.13	0.06	.80	–0.40	.09
Maternal self-efficacy	0.24	.36	–0.27	.31	0.47	.06

[a]For each comparison, the first group is the reference group.

[b]MMB: MumMoodBooster.

[c]TAU: treatment as usual.

[d]FTF-CBT: face-to-face cognitive behavioral therapy.

[e]Significant effects at $P<.05$.

[f]CBT: cognitive behavioral therapy.

Use of Other Treatments

Of the 79.3% (92/116) of participants who responded to this questionnaire, approximately 19% (6/32) of participants in the TAU group and 6% (2/31) of participants in the FTF-CBT group reported using depression medication. None of the MMB participants reported using medication. Approximately 3% (1/31) of women in the FTF-CBT group, 7% (2/29) of women in the MMB group, and 6% (2/32) of women in the TAU group reported having taken part in group therapy. Approximately 23% (7/31), 21% (6/29), and 13% (4/32) of women in the FTF-CBT, MMB, and TAU conditions, respectively, reported using self-help books, and 2 women each in the MMB and TAU groups reported having used acupuncture or hypnotherapy.

Helpfulness and Satisfaction

Participants in the 2 active conditions (FTF-CBT and MMB) rated the helpfulness of their treatment on a rating scale of 0 to 3. Compared with those in the MMB group (mean 2.31, SD 0.89), participant ratings in the FTF-CBT group (mean 2.84, SD 0.38) were significantly higher ($P=.02$).

MMB participants rated the helpfulness of the coach calls an average of 2.56 (SD 0.67) out of a possible 3 points. MMB participants further rated their satisfaction (from 0 to 3) with the following features of the internet program: mood tracking (mean 2.06, SD 0.91), pleasant activities (mean 2.28, SD 0.85), strategies to reduce negative thinking (mean 2.26, SD 1.03), strategies to increase positive thinking (mean 2.19, SD 0.91), partner support program (mean 1.19, SD 1.02), library articles (mean 1.96, SD 1.04), and videos (mean 1.87, SD 1.07).

Discussion

Existing research confirms that PND can be successfully treated by internet-delivered psychological interventions [51]. However, no previously published study has made a direct comparison of the efficacy of such interventions with traditional FTF treatment in a clinically diagnosed population.

Principal Findings

In a sample of postpartum women, 92.2% (107/116) of whom had a major depression diagnosis and were in the moderate-to-severe range for depression symptoms on the BDI-II, we found that the MMB internet-delivered treatment program performed at least as well as FTF-CBT on remission from a diagnosed depressive episode at 21 weeks. Moreover, MMB was also significantly more effective than FTF-CBT in reducing depression symptom severity and perceived stress over time. Compared with TAU, MMB was significantly more effective in reducing depression, anxiety, and perceived stress. At follow-up, depression and anxiety symptom scores for MMB participants were approximately 50% lower than those observed in the FTF-CBT and TAU conditions.

These results are broadly consistent with the literature that delivery of psychological treatments via the internet can be as effective as traditional FTF approaches for treating depression and anxiety [26-28,88,89]. Moderate-to-large effects are generally reported [90], although it should be noted that in many studies, waitlist controls and TAU are the typical comparators. The findings also parallel the results from our previous RCT [43], which showed a significantly larger reduction in depression symptoms in the MMB condition compared with TAU.

The use of coaches as an adjunct to the MMB program may have contributed to the relatively low attrition in this condition and enhanced the effectiveness of the MMB program. This is consistent with our previous trials of MMB [43,91] and with current evidence suggesting that fully automated treatments can be enhanced through support, whether guided by a technician or therapist via phone, email, or SMS text messages [19,29,31]. For example, Karyotaki et al [92] conducted a meta-analysis of a range of guided internet interventions for depression and found

significantly better efficacy for guided interventions compared with fully automated interventions. The current evidence base confirms that self-guided interventions do produce benefits; however, guided support seems to enhance the efficacy of many internet interventions [25,27,93].

Strengths

This study has a number of strengths. First, this RCT produced very similar effect sizes for reductions in depression and anxiety symptoms and the rate of session attendance and program adherence compared with those we observed in our 2 previous trials of MMB [42,43]. In our earlier RCT [43], MMB produced a 4-fold difference in remission compared with TAU as measured at the 12-week follow-up. In this study, follow-up occurred considerably later (at 21 weeks). Remission rates in the MMB condition were still found to be similar to those in our previous RCT [43] and superior to TAU, albeit with a smaller between-group difference in depression remission rate than at 12 weeks in the previous trial. This may indicate that the gains made during the program were maintained over this longer follow-up period. This difference between our previous RCT (measured at 12 weeks) and in our current RCT (measured at 21 weeks) is perhaps not unexpected, given that meta-analyses suggest considerable spontaneous remission rates even in untreated individuals over longer time frames [7].

Importantly, this research provides the first direct comparison of internet-delivered psychological treatment for perinatal depression with evidence-based FTF treatment. This was achieved in a sample of women clinically diagnosed with depressive disorders rather than in a sample of women with a mix of subclinical symptoms assessed using screening tools as in many existing studies.

Furthermore, the rate at which participants accessed other treatments outside of those provided in the trial was low and appeared similar in the 2 active treatment groups. However, despite the higher use of *other supports* found in the TAU group, the participants assigned to the MMB condition reported greater reductions in depression and anxiety symptom scores.

Our use of a 3-arm RCT that included TAU and intention-to-treat analyses gives us further confidence in interpreting these results. As such, this study extends the literature showing promising results from trials of internet-based treatments for PND [50,51].

The FTF-CBT intervention was primarily provided by practicing psychologists in the community, making the comparison as close an approximation as possible to a real-world treatment option. In Australia, FTF-CBT is 1 of the 2 main treatments that postnatal women with depression are likely to be offered (the other being antidepressant medication).

As our results showed that MMB was at least comparable with traditional FTF-CBT in efficacy, its more widespread implementation could potentially increase the reach of treatment to benefit women with PND not currently accessing traditional services because of geographic or social isolation or through a preference for anonymous, nonclinical settings.

Limitations

The study has a number of limitations that should be noted, some of which are common to most RCTs. Our observed 21-week outcomes may have reflected some cases in which relapse occurred after treatment completion, perhaps in both of the active treatments as well as some spontaneous recovery over time in the TAU group. Participants in this study agreed to be assigned to any of the 3 possible experimental conditions; therefore, they may be a particular subset of the population and not representative of postnatal women with depression as a whole. It is possible that our study results may not generalize to women who would have preferred to use only the internet intervention (MMB) or alternatively only attend the clinic-based FTF-CBT treatment. Furthermore, given the nature of the intervention, the participants could not be blinded beyond the point of treatment allocation. Finally, the limited sample size was not sufficiently powered to examine subgroup differences or the moderating effects of baseline participant characteristics.

Future Directions

Emerging evidence suggests that the COVID-19 pandemic is seeing substantial increases in perinatal mental health difficulties [94,95] concurrent with mandatory restrictions on population movement. Therefore, access to FTF psychological services may be problematic in many locations, and effective, accessible internet interventions for perinatal depression are likely to be highly relevant. There is increasing recognition that the impact of real-world implementation of internet mental health interventions and their cost-effectiveness [96] depends upon how treatments are configured and integrated into the mental health treatment landscape as a whole [97,98].

Various options for deploying digital delivery as part of a *blended* or guided approach that combines clinical therapy or counseling with internet program content show promise [99-101], with relevance across a range of diagnoses [102]. Systematic reviews suggest that programs that combine traditional and digital approaches can achieve clinically meaningful effects on depression and anxiety outcomes [103], and based on current evidence [96,104,105], it seems reasonable to infer that internet CBT could also be a cost-effective complement to the support provided to women with PND in primary care. An important potential advantage is that internet CBT can be commenced more or less immediately, with reduced or no waiting time. This could facilitate very early intervention if, for example, the intervention program were linked to routine depression screening in the very early postpartum period.

Evaluation is also warranted for the stepped-care model or adaptive treatment strategies for moderate depression and anxiety symptoms that offer a standalone internet intervention as an initial treatment option [106,107]. Future research is needed to examine program effectiveness and implementation scenarios associated with offering standalone internet PND interventions that do not include human guidance. For example, the scalability and sustainability of supported internet CBT depend at least in part on the resources available to such programs, so that unguided programs of comparable efficacy may be a useful alternative in many service delivery contexts [34]. Important questions also remain around issues such as

patient adherence and the specific health care population from which users of internet psychological interventions are drawn, both in the real-world setting [108,109] and in efficacy RCTs [110]. We found some evidence that session attendance in the MMB condition may have been somewhat higher than in FTF-CBT. Finally, variations of (or complements to) MMB deserve additional research attention, including internet depression treatment programs for pregnant women as well as for new fathers who have so far been largely neglected.

Conclusions

The current trial replicated key findings of prior MMB studies, providing further evidence that internet-delivered treatment for postnatal women with depression is a viable treatment option. These findings are broadly consistent with the notion that speed of access, anonymity, and convenience may be specific advantages of internet-delivered treatment for women who prefer not to engage in traditional treatment modalities. Notably, this study also provides empirical evidence that participants assigned to the MMB condition report greater treatment gains compared with those assigned to the TAU and FTF-CBT conditions.

Acknowledgments

The authors thank all the women who took part in the study. Sarah Maher at the Parent-Infant Research Institute kindly assisted with data retrieval, management, screening, and cleaning. This work was supported by a grant from the National Health and Medical Research Council, Australia (project grant 1046875).

Authors' Contributions

JM, BGD, JRS, MST, JE, and CH conceived the MumMoodBooster intervention. JM, AWG, BGD, JRS, CH, and CJH contributed to the study design and funding applications. JM, CJH, CH, JE, and AWG implemented the trial procedures and collected the data. CH, JM, and JE trained the phone coaches. CJH project managed this trial. JRS (50%), JM (30%), BGD (10%), and AWG (10%) designed and executed the statistical analysis. AWG and JRS wrote the bulk of the first draft of the manuscript. All authors contributed to subsequent drafts, interpretation of results, and approval of the final manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT e-HEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2462 KB - jmir_v23i12e17185_app1.pdf]

References

1. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
2. Beck CT. The effects of postpartum depression on child development: a meta-analysis. Arch Psychiatr Nurs 1998 Feb;12(1):12-20. [doi: 10.1016/s0883-9417(98)80004-6] [Medline: 9489170]
3. Cornish A, McMahon C, Ungerer J, Barnett B, Kowalenko N, Tennant C. Postnatal depression and infant cognitive and motor development in the second postnatal year: the impact of depression chronicity and infant gender. Infant Behav Develop 2005 Dec;28(4):407-417. [doi: 10.1016/j.infbeh.2005.03.004]
4. Milgrom J, Westley DT, Gemmill AW. The mediating role of maternal responsiveness in some longer term effects of postnatal depression on infant development. Infant Behav Develop 2004 Dec;27(4):443-454. [doi: 10.1016/j.infbeh.2004.03.003]
5. Kaplan PS, Danko CM, Diaz A, Kalinka CJ. An associative learning deficit in 1-year-old infants of depressed mothers: role of depression duration. Infant Behav Dev 2011 Feb;34(1):35-44 [FREE Full text] [doi: 10.1016/j.infbeh.2010.07.014] [Medline: 21071090]
6. Pawlby S, Sharp D, Hay D, O'Keane V. Postnatal depression and child outcome at 11 years: the importance of accurate diagnosis. J Affect Disord 2008 Apr;107(1-3):241-245. [doi: 10.1016/j.jad.2007.08.002] [Medline: 17854906]
7. Whiteford HA, Harris MG, McKeon G, Baxter A, Pennell C, Barendregt JJ, et al. Estimating remission from untreated major depression: a systematic review and meta-analysis. Psychol Med 2013 Aug;43(8):1569-1585. [doi: 10.1017/S0033291712001717] [Medline: 22883473]
8. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. J Affect Disord 2014 Apr;159:118-126. [doi: 10.1016/j.jad.2014.02.026] [Medline: 24679399]

9. National Institute of Mental Health. Breaking ground, breaking through: The strategic plan for mood disorders research. USA: National Institute of Health; 2003.
10. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16-e50 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
11. Dennis C, Hodnett E. Psychosocial and psychological interventions for treating postpartum depression. Cochrane Database Syst Rev 2007 Oct 17(4):CD006116. [doi: 10.1002/14651858.CD006116.pub2] [Medline: 17943888]
12. Cuijpers P, Brännmark JG, van Straten A. Psychological treatment of postpartum depression: a meta-analysis. J Clin Psychol 2008 Jan;64(1):103-118. [doi: 10.1002/jclp.20432] [Medline: 18161036]
13. Milgrom J, Negri LM, Gemmill AW, McNeil M, Martin PR. A randomized controlled trial of psychological interventions for postnatal depression. Br J Clin Psychol 2005 Nov;44(Pt 4):529-542. [doi: 10.1348/014466505X34200] [Medline: 16368032]
14. MacArthur C, Winter H, Bick D, Knowles H, Lilford R, Henderson C, et al. Effects of redesigned community postnatal care on womens' health 4 months after birth: a cluster randomised controlled trial. Lancet 2002 Feb 02;359(9304):378-385. [doi: 10.1016/s0140-6736(02)07596-7] [Medline: 11844507]
15. Appleby L, Warner R, Whitton A, Faragher B. A controlled study of fluoxetine and cognitive-behavioural counselling in the treatment of postnatal depression. BMJ 1997 Mar 29;314(7085):932-936 [FREE Full text] [doi: 10.1136/bmj.314.7085.932] [Medline: 9099116]
16. O'Hara MW, Stuart S, Gorman LL, Wenzel A. Efficacy of interpersonal psychotherapy for postpartum depression. Arch Gen Psychiatry 2000 Nov 01;57(11):1039-1045. [doi: 10.1001/archpsyc.57.11.1039] [Medline: 11074869]
17. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
18. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec 15;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
19. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
20. 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]
21. Carroll KM, Ball SA, Martino S, Nich C, Babuscio TA, Rounsaville BJ. Enduring effects of a computer-assisted training program for cognitive behavioral therapy: a 6-month follow-up of CBT4CBT. Drug Alcohol Depend 2009 Feb 01;100(1-2):178-181 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.09.015] [Medline: 19041197]
22. Knaevelsrud C, Maercker A. Internet-based treatment for PTSD reduces distress and facilitates the development of a strong therapeutic alliance: a randomized controlled clinical trial. BMC Psychiatry 2007 Apr 19;7:13 [FREE Full text] [doi: 10.1186/1471-244X-7-13] [Medline: 17442125]
23. Preschl B, Maercker A, Wagner B. The working alliance in a randomized controlled trial comparing online with face-to-face cognitive-behavioral therapy for depression. BMC Psychiatry 2011 Dec 06;11:189 [FREE Full text] [doi: 10.1186/1471-244X-11-189] [Medline: 22145768]
24. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
25. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb 02;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
26. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
27. 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;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
28. Johansson O, Bjärehed J, Andersson G, Carlbring P, Lundh L. Effectiveness of guided internet-delivered cognitive behavior therapy for depression in routine psychiatry: a randomized controlled trial. Internet Interv 2019 Sep;17:100247 [FREE Full text] [doi: 10.1016/j.invent.2019.100247] [Medline: 31249791]
29. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. Aust N Z J Psychiatry 2009 Jun 01;43(6):571-578. [doi: 10.1080/00048670902873722] [Medline: 19440890]
30. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]

31. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 08;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
32. Wiersma JE, Titov N, Van Schaik DJ, Van Oppen P, Beekman AT, Cuijpers P, et al. Treating chronic symptoms of depression in the virtual clinic: findings on chronicity of depression in patients treated with internet-based computerized cognitive behaviour therapy for depression. Psychother Psychosom 2011;80(5):313-315. [doi: 10.1159/000323167] [Medline: 21720195]
33. Andrews G, Williams AD. Up-scaling clinician assisted internet cognitive behavioural therapy (iCBT) for depression: a model for dissemination into primary care. Clin Psychol Rev 2015 Nov;41:40-48. [doi: 10.1016/j.cpr.2014.05.006] [Medline: 25043445]
34. Health Quality Ontario. Internet-delivered cognitive behavioural therapy for major depression and anxiety disorders: a health technology assessment. Ont Health Technol Assess Ser 2019;19(6):1-199 [FREE Full text] [Medline: 30873251]
35. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 07;192(S11):S4-11. [doi: 10.5694/j.1326-5377.2010.tb03685.x] [Medline: 20528707]
36. Andrews G, Cuijpers P, Craske MG, 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]
37. Austin M, Frilingos M, Lumley J, Hadzi-Pavlovic D, Roncolato W, Acland S, et al. Brief antenatal cognitive behaviour therapy group intervention for the prevention of postnatal depression and anxiety: a randomised controlled trial. J Affect Disord 2008 Jan;105(1-3):35-44. [doi: 10.1016/j.jad.2007.04.001] [Medline: 17490753]
38. Gentile S. Use of contemporary antidepressants during breastfeeding: a proposal for a specific safety index. Drug Saf 2007;30(2):107-121. [doi: 10.2165/00002018-200730020-00002] [Medline: 17253877]
39. Bilszta J, Ericksen J, Buist A, Milgrom J. Women's experiences of postnatal depression - beliefs and attitudes as barriers to care. Australian J Advanced Nurs 2010;27(3):44-54 [FREE Full text]
40. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
41. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: formative research and design of the MomMoodBooster program. JMIR Res Protoc 2012 Nov 22;1(2):e18 [FREE Full text] [doi: 10.2196/resprot.2329] [Medline: 23612274]
42. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
43. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
44. O'Mahen H, Himle JA, Fedock G, Henshaw E, Flynn H. A pilot randomized controlled trial of cognitive behavioral therapy for perinatal depression adapted for women with low incomes. Depress Anxiety 2013 Jul 14;30(7):679-687. [doi: 10.1002/da.22050] [Medline: 23319454]
45. O'Mahen HA, Woodford J, McGinley J, Warren FC, Richards DA, Lynch TR, et al. Internet-based behavioral activation--treatment for postnatal depression (Netmums): a randomized controlled trial. J Affect Disord 2013 Sep 25;150(3):814-822. [doi: 10.1016/j.jad.2013.03.005] [Medline: 23602514]
46. Pugh NE, Hadjistavropoulos HD, Dirkse D. A randomised controlled trial of therapist-assisted, internet-delivered cognitive behavior therapy for women with maternal depression. PLoS One 2016 Mar 1;11(3):e0149186 [FREE Full text] [doi: 10.1371/journal.pone.0149186] [Medline: 26930488]
47. O'Mahen HA, Grieve H, Jones J, McGinley J, Woodford J, Wilkinson EL. Women's experiences of factors affecting treatment engagement and adherence in internet delivered behavioural activation for postnatal depression. Internet Interventions 2015 Mar;2(1):84-90. [doi: 10.1016/j.invent.2014.11.003]
48. Loughnan SA, Butler C, Sie AA, Grierson AB, Chen AZ, Hobbs MJ, et al. A randomised controlled trial of 'MUMentum postnatal': internet-delivered cognitive behavioural therapy for anxiety and depression in postpartum women. Behav Res Ther 2019 May;116:94-103. [doi: 10.1016/j.brat.2019.03.001] [Medline: 30877878]
49. Loughnan SA, Joubert AE, Grierson A, Andrews G, Newby JM. Internet-delivered psychological interventions for clinical anxiety and depression in perinatal women: a systematic review and meta-analysis. Arch Womens Ment Health 2019 Dec 17;22(6):737-750. [doi: 10.1007/s00737-019-00961-9] [Medline: 31101993]
50. Lee EW, Denison FC, Hor K, Reynolds RM. Web-based interventions for prevention and treatment of perinatal mood disorders: a systematic review. BMC Pregnancy Childbirth 2016 Feb 29;16:38 [FREE Full text] [doi: 10.1186/s12884-016-0831-1] [Medline: 26928898]
51. Ashford MT, Olander EK, Ayers S. Computer- or web-based interventions for perinatal mental health: a systematic review. J Affect Disord 2016 Jun;197:134-146. [doi: 10.1016/j.jad.2016.02.057] [Medline: 26991368]

52. Lau Y, Htun TP, Wong SN, Tam WS, Klainin-Yobas P. Therapist-supported internet-based cognitive behavior therapy for stress, anxiety, and depressive symptoms among postpartum women: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 28;19(4):e138 [FREE Full text] [doi: 10.2196/jmir.6712] [Medline: 28455276]
53. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
54. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
55. Cox J, Holden J. Perinatal Mental Health. A Guide to the Edinburgh Postnatal Depression Scale (EPDS). London: Gaskell; 2003.
56. First M, Spitzer R, Gibbon M. Structured Clinical Interview for DSM-IV Axis I Disorders, Patient Edition (SCID-I/P, Version 2.0). Washington DC: American Psychiatric Press; 1996.
57. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. J Med Internet Res 2005 May 18;7(2):e12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
58. Musiat P, Hoffmann L, Schmidt U. Personalised computerised feedback in E-mental health. J Ment Health 2012 Aug 08;21(4):346-354. [doi: 10.3109/09638237.2011.648347] [Medline: 22315961]
59. Paulson JF, Bazemore SD. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. JAMA 2010 May 19;303(19):1961-1969. [doi: 10.1001/jama.2010.605] [Medline: 20483973]
60. Solness CL, Kroska EB, Holdefer PJ, O'Hara MW. Treating postpartum depression in rural veterans using internet delivered CBT: program evaluation of MomMoodBooster. J Behav Med 2021 Aug 14;44(4):454-466 [FREE Full text] [doi: 10.1007/s10865-020-00188-5] [Medline: 33052526]
61. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. Web-based intervention for postpartum depression: formative research and design of the MomMoodBooster program. JMIR Res Protoc 2012 Nov 22;1(2):e18 [FREE Full text] [doi: 10.2196/resprot.2329] [Medline: 23612274]
62. Milgrom J, Martin P, Negri L. Treating Postnatal Depression. A Psychological Approach for Health Care Practitioners. Chichester: Wiley; 1999.
63. Milgrom J, Gemmill AW, Ericksen J, Burrows G, Buist A, Reece J. Treatment of postnatal depression with cognitive behavioural therapy, sertraline and combination therapy: a randomised controlled trial. Aust N Z J Psychiatry 2015 Mar 13;49(3):236-245. [doi: 10.1177/0004867414565474] [Medline: 25586754]
64. Milgrom J, Holt CJ, Gemmill AW, Ericksen J, Leigh B, Buist A, et al. Treating postnatal depressive symptoms in primary care: a randomised controlled trial of GP management, with and without adjunctive counselling. BMC Psychiatry 2011 May 27;11:95 [FREE Full text] [doi: 10.1186/1471-244X-11-95] [Medline: 21615968]
65. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders - Text Revision. 4th edition. Washington, DC: American Psychiatric Press; 2000.
66. Beck A, Steer R, Brown G. Manual for the BDI-II. San Antonio, TX: The Psychological Corporation; 1996.
67. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. Clin Psychol Rev 1988 Jan;8(1):77-100. [doi: 10.1016/0272-7358(88)90050-5]
68. Beck CT, Gable RK. Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. Nurs Res 2001;50(4):242-250. [doi: 10.1097/00006199-200107000-00008] [Medline: 11480533]
69. Milgrom J, Ericksen J, Negri L, Gemmill AW. Screening for postnatal depression in routine primary care: properties of the Edinburgh Postnatal Depression Scale in an Australian sample. Aust N Z J Psychiatry 2005 Sep 26;39(9):833-839. [doi: 10.1080/j.1440-1614.2005.01660.x] [Medline: 16168042]
70. Ji S, Long Q, Newport DJ, Na H, Knight B, Zach EB, et al. Validity of depression rating scales during pregnancy and the postpartum period: impact of trimester and parity. J Psychiatr Res 2011 Feb;45(2):213-219 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.05.017] [Medline: 20542520]
71. 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]
72. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
73. Sidebottom AC, Harrison PA, Godecker A, Kim H. Validation of the Patient Health Questionnaire (PHQ)-9 for prenatal depression screening. Arch Womens Ment Health 2012 Oct;15(5):367-374. [doi: 10.1007/s00737-012-0295-x] [Medline: 22983357]
74. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
75. Rief W, Nanke A, Klaiberg A, Braehler E. Base rates for panic and depression according to the Brief Patient Health Questionnaire: a population-based study. J Affect Disord 2004 Oct 15;82(2):271-276. [doi: 10.1016/j.jad.2003.11.006] [Medline: 15488257]

76. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
77. 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] [Medline: 7726811]
78. Lovibond S, Lovibond P. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation; 1996.
79. Hollon SD, Kendall PC. Cognitive self-statements in depression: development of an automatic thoughts questionnaire. Cogn Ther Res 1980 Dec;4(4):383-395. [doi: 10.1007/bf01178214]
80. Kanter JW, Mulick PS, Busch AM, Berlin KS, Martell CR. The Behavioral Activation for Depression Scale (BADS): psychometric properties and factor structure. J Psychopathol Behav Assess 2006 Oct 25;29(3):191-202. [doi: 10.1007/s10862-006-9038-5]
81. Wittkowski A, Garrett C, Calam R, Weisberg D. Self-report measures of parental self-efficacy: a systematic review of the current literature. J Child Fam Stud 2017 Jul 6;26(11):2960-2978 [FREE Full text] [doi: 10.1007/s10826-017-0830-5] [Medline: 29081640]
82. Johnston C, Mash EJ. A measure of parenting satisfaction and efficacy. J Clin Child Psychol 1989 Jun;18(2):167-175. [doi: 10.1207/s15374424jccp1802_8]
83. Hunsley M, Best M, Lefebvre D, Vito J. The seven-item short form of the dyadic adjustment scale: further evidence for construct validity. Am J Family Ther 2001 Oct;29(4):325-335. [doi: 10.1080/01926180126501]
84. Simon GE, Ludman EJ, Tutty S, Operskalski B, Von Korff M. Telephone psychotherapy and telephone care management for primary care patients starting antidepressant treatment: a randomized controlled trial. JAMA 2004 Aug 25;292(8):935-942. [doi: 10.1001/jama.292.8.935] [Medline: 15328325]
85. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991 Sep 01;48(9):851-855. [doi: 10.1001/archpsyc.1991.01810330075011] [Medline: 1929776]
86. Feingold A. A regression framework for effect size assessments in longitudinal modeling of group differences. Rev Gen Psychol 2013 Mar 01;17(1):111-121 [FREE Full text] [doi: 10.1037/a0030048] [Medline: 23956615]
87. Zhang Z, Wang L. Statistical power analysis for growth curve models using SAS. Behav Res Methods 2009 Nov;41(4):1083-1094. [doi: 10.3758/BRM.41.4.1083] [Medline: 19897816]
88. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
89. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
90. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
91. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
92. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
93. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
94. Berthelot N, Lemieux R, Garon-Bissonnette J, Drouin-Maziade C, Martel, Maziade M. Uptrend in distress and psychiatric symptomatology in pregnant women during the coronavirus disease 2019 pandemic. Acta Obstet Gynecol Scand 2020 Jul;99(7):848-855. [doi: 10.1111/aogs.13925] [Medline: 32449178]
95. Cameron EE, Joyce KM, Delaquis CP, Reynolds K, Protudjer JL, Roos LE. Maternal psychological distress and mental health service use during the COVID-19 pandemic. J Affect Disord 2020 Nov 01;276:765-774 [FREE Full text] [doi: 10.1016/j.jad.2020.07.081] [Medline: 32736186]
96. Kolovos S, van DJ, Riper H, Buntrock C, Cuijpers P, Ebert DD, et al. Cost effectiveness of guided internet-based interventions for depression in comparison with control conditions: an individual-participant data meta-analysis. Depress Anxiety 2018 Jan 12;35:209-219. [doi: 10.1002/da.22714] [Medline: 29329486]

97. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
98. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Dec 03;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
99. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the Efficiency of Psychotherapy for Depression: Computer-Assisted Versus Standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250. [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
100. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
101. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
102. Weisel KK, Zarski A, Berger T, Schaub MP, Krieger T, Moser CT, et al. Transdiagnostic tailored internet- and mobile-based guided treatment for major depressive disorder and comorbid anxiety: study protocol of a randomized controlled trial. Front Psychiatry 2018 Jul 4;9:274 [FREE Full text] [doi: 10.3389/fpsyt.2018.00274] [Medline: 30022954]
103. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
104. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
105. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010 Dec 19;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
106. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
107. Health Quality Ontario. Internet-delivered cognitive behavioural therapy for major depression and anxiety disorders: a health technology assessment. Ont Health Technol Assess Ser 2019;19(6):1-199 [FREE Full text] [Medline: 30873251]
108. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
109. Romijn G, Batelaan N, Kok R, Koning J, van Balkom A, Titov N, et al. Internet-delivered cognitive behavioral therapy for anxiety disorders in open community versus clinical service recruitment: meta-analysis. J Med Internet Res 2019 Apr 17;21(4):e11706 [FREE Full text] [doi: 10.2196/11706] [Medline: 30994462]
110. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug 6;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]

Abbreviations

ATQ: Automatic Thoughts Questionnaire
BDI-II: Beck Depression Inventory, revised
CBT: cognitive behavioral therapy
DASS-21: Depression Anxiety Stress Scale, 21-item
***DSM-IV*:** *Diagnostic and Statistical Manual for Mental Disorders, Fourth Edition*
FTF: face-to-face
GP: general practitioner
MMB: MumMoodBooster
PHQ-9: Patient Health Questionnaire, 9-item
PND: postnatal depression
RCT: randomized controlled trial
SCID-IV: Structured Clinical Interview for the DSM-IV
TAU: treatment as usual

Edited by R Kukafka, G Eysenbach; submitted 18.11.20; peer-reviewed by E Smoktunowicz, K Champion, A Etzelmueller; comments to author 18.12.20; revised version received 10.03.21; accepted 28.09.21; published 08.12.21.

Please cite as:
Milgrom J, Danaher BG, Seeley JR, Holt CJ, Holt C, Ericksen J, Tyler MS, Gau JM, Gemmill AW
Internet and Face-to-face Cognitive Behavioral Therapy for Postnatal Depression Compared With Treatment as Usual: Randomized Controlled Trial of MumMoodBooster
J Med Internet Res 2021;23(12):e17185
URL: https://www.jmir.org/2021/12/e17185
doi:10.2196/17185
PMID:34889742

Original Paper

Long-Term Effects of a Web-Based Low-FODMAP Diet Versus Probiotic Treatment for Irritable Bowel Syndrome, Including Shotgun Analyses of Microbiota: Randomized, Double-Crossover Clinical Trial

Dorit Vedel Ankersen[1], BSc, MSc, PhD; Petra Weimers[1], MD, PhD; Mette Bennedsen[1], MD, PhD; Anne Birgitte Haaber, MD; Eva Lund Fjordside[1], MD; Moritz Emanuel Beber[2], PhD; Christian Lieven[2], PhD; Sanaz Saboori[1], MD; Nicolai Vad[1], MD; Terje Rannem[1], MD; Dorte Marker[1], RN; Kristine Paridaens[3], MD; Suzanne Frahm[1], RD; Lisbeth Jensen[4], RD; Malte Rosager Hansen[1], MD; Johan Burisch[1], MD, PhD; Pia Munkholm[1], MDSci

[1]Department of Gastroenterology, North Zealand University Hospital, Frederikssund, Denmark

[2]Unseen Bio Anpartsselskab, Copenhagen, Denmark

[3]Ferring International Center Société Anonyme, Saint-Prex, Switzerland

[4]Department of Dietetics, Herlev University Hospital, Herlev, Denmark

Corresponding Author:
Pia Munkholm, MDSci
Department of Gastroenterology
North Zealand University Hospital
Frederikssundsvej 30
Frederikssund, 3600
Denmark
Phone: 45 48292078
Fax: 45 48293611
Email: pia.munkholm@regionh.dk

Abstract

Background: The long-term management of irritable bowel syndrome (IBS) poses many challenges. In short-term studies, eHealth interventions have been demonstrated to be safe and practical for at-home monitoring of the effects of probiotic treatments and a diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs). IBS has been linked to alterations in the microbiota.

Objective: The aim of this study was to determine whether a web-based low-FODMAP diet (LFD) intervention and probiotic treatment were equally good at reducing IBS symptoms, and whether the response to treatments could be explained by patients' microbiota.

Methods: Adult IBS patients were enrolled in an open-label, randomized crossover trial (for nonresponders) with 1 year of follow-up using the web application IBS Constant Care (IBS CC). Patients were recruited from the outpatient clinic at the Department of Gastroenterology, North Zealand University Hospital, Denmark. Patients received either VSL#3 for 4 weeks (2 × 450 billion colony-forming units per day) or were placed on an LFD for 4 weeks. Patients responding to the LFD were reintroduced to foods high in FODMAPs, and probiotic responders received treatments whenever they experienced a flare-up of symptoms. Treatment response and symptom flare-ups were defined as a reduction or increase, respectively, of at least 50 points on the IBS Severity Scoring System (IBS-SSS). Web-based ward rounds were performed daily by the study investigator. Fecal microbiota were analyzed by shotgun metagenomic sequencing (at least 10 million 2 × 100 bp paired-end sequencing reads per sample).

Results: A total of 34 IBS patients without comorbidities and 6 healthy controls were enrolled in the study. Taken from participating subjects, 180 fecal samples were analyzed for their microbiota composition. Out of 21 IBS patients, 12 (57%) responded to the LFD and 8 (38%) completed the reintroduction of FODMAPs. Out of 21 patients, 13 (62%) responded to their first treatment of VSL#3 and 7 (33%) responded to multiple VSL#3 treatments. A median of 3 (IQR 2.25-3.75) probiotic treatments were needed for sustained symptom control. LFD responders were reintroduced to a median of 14.50 (IQR 7.25-21.75)

high-FODMAP items. No significant difference in the median reduction of IBS-SSS for LFD versus probiotic responders was observed, where for LFD it was –126.50 (IQR –196.75 to –76.75) and for VSL#3 it was –130.00 (IQR –211.00 to –70.50; *P*>.99). Responses to either of the two treatments were not able to be predicted using patients' microbiota.

Conclusions: The web-based LFD intervention and probiotic treatment were equally efficacious in managing IBS symptoms. The response to treatments could not be explained by the composition of the microbiota. The IBS CC web application was shown to be practical, safe, and useful for clinical decision making in the long-term management of IBS. Although this study was underpowered, findings from this study warrant further research in a larger sample of patients with IBS to confirm these long-term outcomes.

Trial Registration: ClinicalTrials.gov NCT03586622; https://clinicaltrials.gov/ct2/show/NCT03586622

(J Med Internet Res 2021;23(12):e30291) doi:10.2196/30291

KEYWORDS

irritable bowel syndrome; web-based low-FODMAP diet; probiotics; randomized trial; web-based; IBS; symptom management; treatment outcomes; outcomes; treatment; microbiota; microbiome; gastroenterology; mobile app; mHealth; eHealth

Introduction

Irritable bowel syndrome (IBS) is a functional gastrointestinal disorder that affects 10% to 20% of the population in westernized countries [1]. The main gastrointestinal symptoms include bloating, pain, diarrhea, constipation, and altered bowel habits; psychological comorbidity is common with IBS, thereby symptoms are often accompanied by coexisting conditions like stress, anxiety, and depression that, together, greatly impact patients' quality of life (QoL) [1-3]. Studies have shown that approximately 60% [4] of IBS onset is associated with psychosocial stressors, and up to 32% [5] is associated with prior acute gastroenteritis. The pathophysiology of IBS is not fully understood, but some of the underlying mechanisms include altered gastrointestinal motility, visceral hypersensitivity, psychosocial disturbance, low-grade inflammation, altered gut-brain function, and microbial ecosystem dysbiosis [3,6].

Managing IBS continues to be challenging because of the complexity of its chronicity, heterogeneous patient groups with and without comorbidities, and a lack of both diagnostic tools and well-documented treatment strategies for long-term purposes. Treatment strategies for IBS range from lifestyle and dietary advice to pharmacological solutions that generally target only the primary symptom. Recently, holistic eHealth solutions that monitor symptoms and support patients with digestive diseases have proven valuable in involving and empowering patients, and are used as an adjuvant to current treatment regimens [7,8]. Ideally, an app for IBS would include dietary education supported by clinical dieticians, since dietary triggers are reported to be central to symptom generation in 50% to 84% of patients with IBS [7,9].

An exclusionary diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs) has proven effective in reducing gastrointestinal symptoms, in particular abdominal pain, bloating, and diarrhea [10], with a response rate between 50% and 80% among patients with IBS [9]. In addition, a low-FODMAP diet (LFD) has been shown by some studies to improve the disease course of IBS [11,12] and inflammatory bowel disease (IBD) with IBS-like symptoms [11]. A previous study has shown that short-term, web-based management of IBS patients, while treating their symptoms with LFD and probiotics, is practical and effective [13]. Yet, the LFD has also been linked to a reduction in beneficial gastrointestinal bacteria [14] and to compromise nutritional intake, especially calcium [15]. Therefore, eventual reintroduction of foods high in FODMAPs is essential in long-term management [9].

The composition of gut microbiota differs between subgroups of IBS patients and healthy individuals [16]. This fact, together with an increasing awareness of the importance of a "healthy gut" and the gastrointestinal microbiome, has led to growing research, financial investment, and consumer interest in probiotic treatments that may provide benefits to patients with IBS [17,18]. Probiotic treatments are not currently included in recommended treatment strategies for IBS, but in some countries, national guidelines, such as those issued by NICE (the National Institute for Health and Care Excellence) in the United Kingdom [19], suggest that IBS patients could benefit from probiotic treatments for alleviating symptoms. However, the results for various probiotic solutions, as well as their dosing, duration, and efficacy as a treatment for IBS, are inconclusive [18,20,21]. Moreover, probiotic treatments do not seem to sustain any of the microbial changes they induce after cessation of treatment [22].

So far, no one has been able to fingerprint patients based on their fecal microbial composition and tailor a particular intervention for them individually [4].

The primary aim of this 1-year, web-based study was to determine whether a 4-week probiotic treatment, VSL#3 (Actial Farmaceutica Srl), and the LFD were equally as effective at reducing IBS symptoms, and whether the response to either of the two treatments could be explained by fecal microbiota. The study's secondary aims were to evaluate patients' QoL, the effect of multiple VSL#3 treatments in sustaining symptom control across 1 year, and reintroduction of FODMAPs among LFD responders, as well as to evaluate the web application IBS Constant Care (IBS CC) [23] for its efficacy in long-term clinical use.

Methods

Participants

Adult IBS patients, 18 years or older, fulfilling the Rome III criteria and subdiagnosed with either IBS-diarrhea (IBS-D) or IBS–mixed type (IBS-M) by a gastroenterologist were consecutively included in the study at North Zealand University Hospital (August 23, 2018, to October 18, 2019). Patients were excluded if they did not have a smartphone, had previously undergone gastrointestinal surgery, had been diagnosed with celiac disease or lactose intolerance, had been subdiagnosed with constipation or unspecified IBS, were pregnant, followed alternative diets, had a history of alcohol or drug abuse, had a BMI below 18.5 or above 35 (calculated as weight in kg divided by height in m^2), had been diagnosed with any comorbidities (eg, diabetes), had previously been on an LFD guided by a dietician, had been on any probiotic or antibiotic treatment within the 3 months prior to inclusion, or had an IBS severity score lower than 175 at inclusion.

Healthy controls (HCs) older than 18 years, with a normal BMI (18.5-25), fecal calprotectin (FC) less than 70 mg/kg, and not taking daily medication or food supplements were recruited internally from our institution.

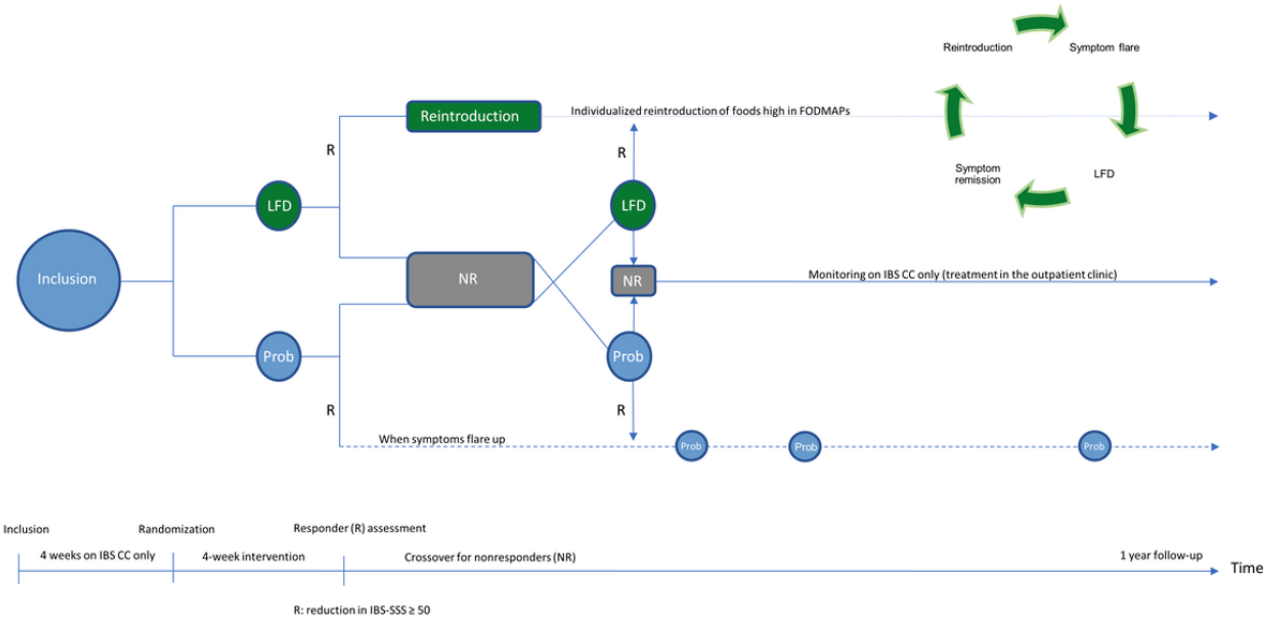
The Ethical Committee of Denmark (H-16023499) and the Danish Data Protection Agency (I-Suite No. 6262, NOH-2018-002) approved this study. The study was registered at ClinicalTrials.gov (NCT03586622; see Multimedia Appendix 1 for the CONSORT-EHEALTH checklist). All patients in the study gave written informed consent prior to their inclusion.

Study Design and Protocol

Overview

This eHealth study was an open-label, randomized, crossover trial (for nonresponders), with 1 year of follow-up (Figure 1). HCs home-monitored their weight, QoL, and symptoms at four points during the year using the IBS CC web application. These measurements were then represented in a traffic light format to the participants (see the Applications Used in the Study section for a description). At inclusion, patients were trained, for approximately 1 hour, in home monitoring of symptoms using IBS CC, including inflammation measured via FC. FC was measured using the CalproSmart app (Calpro AS). Patients were instructed to home-monitor every week for at least 4 weeks before randomization to either the LFD or probiotic treatment group.

Figure 1. The 1-year design of this study. The IBS Constant Care (IBS CC) web application was used for home monitoring of symptoms and clinical decision making. Treatments included 4 weeks of monitoring on IBS CC, followed by randomization to either a 4-week low-FODMAP diet (LFD) or probiotic treatment (Prob; VSL#3, Actial Farmaceutica Srl). Responders (R) to the LFD would subsequently be reintroduced to foods higher in FODMAPs. Responders to probiotic treatment would receive multiple treatments upon symptom flare-ups (ie, an increase in the IBS Severity Scoring System [IBS-SSS] of more than 50 points). Response was defined as a decrease in IBS-SSS of at least 50 points. Nonresponders (NR) waited for a minimum of 2 weeks wash-out before being crossed over (this is not shown in the figure). FODMAPs: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS: irritable bowel syndrome.



Patients randomized to the LFD group had a consultation at the hospital with a nutritionist, who guided the patients individually through the diet's principles and how to monitor their symptoms (see Multimedia Appendix 2 for home monitoring and screening intervals). Patients randomized to the probiotic treatment group were instructed on how to consume the sachets and home-monitor using IBS CC during treatment (Multimedia Appendix 2). After 4 weeks of either the LFD or probiotic treatment, patients' response to treatment was evaluated. Response to treatment was defined as a reduction of at least 50 points in the IBS Severity Scoring System (IBS-SSS) [24]. If patients responded to the LFD, they were scheduled for a consultation on how to reintroduce high-FODMAP foods using IBS CC (described in Multimedia Appendix 2). Patients were instructed to resume a strict LFD for a few days when they encountered any symptom flare-ups (ie, an increase of at least

50 points in their IBS-SSS) during this reintroduction period (Figure 1). If patients responded to probiotic treatment, they were instructed to home-monitor their symptoms every week; if or when they relapsed (ie, saw an increase of at least 50 points in their IBS-SSS), they would receive another 4-week probiotic treatment, and so on, until 1 year of follow-up had elapsed (Figure 1). If patients did not respond to the initial treatment, they were instructed to take a wash-out period of at least 2 weeks, all the while continuing to home-monitor on IBS CC, and were subsequently crossed over. If patients did not respond to either of the two treatments, they could still home-monitor using IBS CC, following treatment advice from the outpatient clinic at the Department of Gastroenterology, North Zealand University Hospital. Web-based ward rounds were performed daily by the study investigator, based on an electronic patient list (green: inactive symptoms; yellow: mild to moderate symptoms; or red: severe symptoms, according to the IBS-SSS). Instructions to patients regarding how often they should monitor themselves at home using the IBS CC web application are listed in Multimedia Appendix 2, and email reminders (ie, "time to home-monitor") were sent to participants that accepted these.

Fecal Sampling

Participants (IBS and HCs) were trained at inclusion in downloading the CalproSmart app and performing the FC home test. Participants received FC kits, including pads, frame, and tubes with buffer fluid, and were instructed to perform an FC home test within the first 4 weeks of home monitoring and again after 1 year of follow-up.

Participants were provided with easy sampler kits and were instructed to either send fecal samples for microbiota analysis via their general practitioner, or to store fecal samples in their domestic freezer (–20 ◦C). Participants (IBS and HCs) delivered their frozen fecal samples to the study investigator whenever they had an on-site consultation at the hospital. All samples were immediately frozen (–80 ◦C) upon receipt at the hospital. Patients were instructed to store or send fecal samples at study inclusion, at randomization, after 4 weeks of follow-up, and after 1 year of follow-up. Patients responding to probiotic treatment were further instructed to store samples every time they received a new 4-week probiotic treatment, sampled before and just after the treatment (Figure 1). HCs were instructed to collect a fecal sample four times during the course of 1 year and to store these samples in their domestic freezer. All fecal samples were shipped on dry ice to the microbiome laboratory in Germany.

Interventions

Low-FODMAP Diet and Reintroduction of High-FODMAP Foods

One hour of dietary advice was provided on an individual basis by an experienced dietician or nutritionist that took into account the dietary history of each patient. Participants received guidance for the LFD using the Danish Low FODMAP Diet app (Muusmann Publisher), and were further instructed to measure their QoL, weight, adherence to the diet, and any symptoms on the IBS CC web application. In addition, participants also received supplementary materials (prepared by the study authors) on the LFD, with suggestions for LFD-friendly meals and recipes (eg, bread and ice cream); these documents were uploaded to each individual patient's folder on IBS CC. Patients were advised to significantly reduce their intake of foods high in FODMAPs. Patients were advised only to consume foods low in FODMAPs, which were labeled green in the Low FODMAP Diet app; only two yellow-labeled foods, in small quantities, were allowed per day for reluctant participants.

Responders in the LFD group were taught to reintroduce foods high in FODMAPs using the food preference approach, all while taking into account their usual, preferred diet (described in Multimedia Appendix 2).

Probiotic Treatment

VSL#3, distributed by Ferring Inc in Europe, is a high-concentration multi-strain probiotic mix containing one strain of *Streptococcus thermophilus* BT01, three strains of Bifidobacteria (*B breve* BB02; *B animalis* subspecies [subsp] *lactis* BL03, previously identified as *B longum* BL03; and *Banimalis* subsp *lactis* BI04, previously identified as *B infantis* BI04), and four strains of Lactobacilli (*L acidophilus* BA05, *L plantarum* BP06, *L paracasei* BP07, and *L helveticus* BD08, previously identified as *L delbrueckii* subsp *bulgaricus* BD08). VSL#3 contains no less than 450 billion colony-forming units (CFU) per sachet. This formulation is sold as an over-the-counter food supplement at pharmacies in Denmark and by a variety of commercial websites.

VSL#3 sachets used in this study were provided by Actial Farmaceutica Srl, Italy, with the following batch numbers: 804033 and 908081. VSL#3 was administered for 26 days (approximately 4 weeks) for subjects allocated to the treatment group. Patients were instructed to take one sachet twice daily and store the sachets in the refrigerator. Patients were advised not to consume VSL#3 together with any hot or carbonated drinks. Side effects, if any, were registered throughout the study period.

Measures

Applications Used in the Study

IBS CC has previously been used in other web-based studies [13,25]. The updated and expanded version of IBS CC [23] used in this study includes various patient-reported outcome measures (PROMs), listed in Multimedia Appendix 2 and described in detail below. Patient-reported satisfaction with the current version of IBS CC was evaluated before randomization and at 1-year follow-up. Participants gained access to IBS CC via a log-in procedure that includes two-factor authentication. IBS CC meets the requirements of the European Union's General Data Protection Regulation.

Participants made their FC measurements at home using the CE (Conformité Européenne)–marked CalproSmart app. This home test can be performed in 18 minutes and is integrated into the IBS CC application, providing patients with an opportunity to see FC and other measures, such as the IBS-SSS, the Bristol Stool Chart (BSC), bowel movement frequency, weight and BMI, and QoL, all longitudinally and in a traffic light form. Two examples of a 1-year disease course, from a responder in

the probiotic group and another in the LFD group, in IBS CC are shown in Multimedia Appendix 2.

The Low FODMAP Diet app is available from the App Store or Google Play and was provided free of charge to the patients in this study. The app includes sections about IBS, the LFD, recipes, grocery lists, a diary, and a module illustrating foods in green, yellow, and red. This mobile app is not integrated with IBS CC.

IBS Severity and Responder Definition

IBS severity was measured using the IBS-SSS [24]. This questionnaire consists of five items (ie, distention, bowel habits, interference with QoL, and two questions about abdominal pain) that are to be rated from 0 to 100 on a visual analog scale (VAS), with a total score ranging from 0 to 500. The IBS-SSS was made available on IBS CC with permission from Mapi Research Trust, Lyon, France. For ease of interpretation and patient involvement with IBS CC, a score of 0 to 175 was classified as "symptom remission and mild activity" (green), 175 to 300 as "moderate activity" (yellow), and 300 or more as "severe activity" (red). These cutoffs have been used in previous IBS web intervention trials [13,25]. Responses to an intervention or a symptom flare-up were defined as a change in IBS-SSS of at least 50 points [24].

Bristol Stool Chart and Bowel Movement Frequency

The BSC illustrates seven different stool types that represent constipation (types 1 and 2), "normal" stools (types 3-5), and diarrhea (types 6 and 7) [9]. Patients were instructed to assess their stool type either daily or weekly, together with bowel movement frequency, using the IBS CC web application. Results were illustrated via IBS CC to the patients longitudinally, with types 1 and 2 and types 6 and 7 in red, and "normal" stool types in green.

Adherence to Treatments

The FODMAP Adherence Report Scale (FARS) and the Medical Adherence Report Scale (MARS) measured adherence to the diet and probiotic treatment, respectively. The FARS and the MARS are self-assessment questionnaires consisting of five questions that have been used previously in IBD web intervention trials [26,27], one IBS study [11], and one IBD-IBS retrospective study [11]. The total scores range from 0 to 25, where 25 indicates maximum adherence. A score of 22 or higher was considered as adherent and was indicated as green to the patients in IBS CC.

Quality of Life

QoL was measured using the IBS-QoL questionnaire [28] (courtesy of Mapi Research Trust, Lyon, France) via IBS CC. The IBS-QoL questionnaire consists of 34 items, each with a 5-point response scale, resulting in a maximum score of 170. Scores were transformed to a scale of 0 to 100, where scores of 0 to 49 were indicated with a red color and a score of 50 or above was indicated in green, meaning a better QoL.

Disease Course Types

The Copenhagen disease course types describe four different IBS disease courses:

- A: Mild IBS with indolent course
- B: Mild IBS with aggressive course
- C: Chronic IBS with continuous course
- D: Chronic IBS with intermittent course.

These IBS types have previously been used in IBS studies by Maagaard et al [11] and Weynants et al [12]. Our patients were instructed to choose the disease course type that best described their course at inclusion and at 1-year follow-up. These measures were accessible to the participants on IBS CC.

Other Measures Included in IBS CC

Patients also registered their weight in kilograms on a weekly basis at home, and this measure was presented to patients longitudinally, together with their BMI, in IBS CC. Patients were instructed to use their bathroom scale, either in the morning or evening, with or without clothes, as long as they did so consistently throughout the study. The number of probiotic treatments needed for sustaining symptom control and the time between probiotic treatments (in days) were evaluated during the course of 1 year for multiple responders. Participants evaluated their satisfaction with IBS CC, including FC home testing, before their randomization on a VAS from 1 to 10, where 10 was the greatest possible satisfaction. Patient satisfaction with the study and IBS CC at 1 year, including FC home testing, was assessed using an e-questionnaire in IBS CC that was prepared by the authors, consisting of nine "yes or no" questions and two other questions: one regarding the time used for home monitoring and the other asking for suggestions on improving IBS CC for future clinical use. Clinical metadata were exported from the IBS CC database for statistical analyses.

Microbiome Analysis

Fecal samples were analyzed by CeGaT GmbH in Tübingen, Germany. The method used by CeGaT GmbH is based on culture-independent, whole-genome, shotgun metagenomic, next-generation sequencing. Genomic DNA was extracted from the participants' fecal samples using CeGaT GmbH's proprietary protocol, including enzymatic cleavage of genomic DNA. Approximately 100 ng of DNA was used per sample for library preparation, which was performed with the Nextera DNA Flex Library Prep Kit (Illumina). Sequencing was performed on the NovaSeq 6000 Sequencing System (Illumina), which resulted in at least 10 million 2 × 100 bp paired-end sequencing reads (2 Gb) per sample. Demultiplexing of the sequencing reads was performed with bcl2fastq Conversion Software (version 2.20; Illumina) [29]. Adapters were trimmed with Skewer [30]. Taxonomic classification and sequence abundance were estimated by Unseen Bio anpartsselskab, Greater Copenhagen, using an in-house bioinformatics pipeline [31] consisting of quality control with FastQC [32] and fastp [33], removal of human DNA by mapping reads with Kraken 2 [34] to its human library, and a summary using MultiQC [35]. The remaining sequencing reads were then mapped to a human gut microbiome-specific reference database, MGnify [36], with Kraken 2 in order to estimate bacterial and archaeal sequence abundance. Sequence abundance counts were then re-estimated to the species level using Bracken [37].

Statistical Analyses

Counts, means, medians, and IQRs were computed from the PROMs. Wilcoxon rank-sum tests, Wilcoxon signed-rank tests with continuity correction, or Dunn all-pair tests were performed to compare medians. Analysis of sequence counts or reads and visualization of results were performed in R (version 4.0.4; The R Foundation). A list of R packages used in this study can be found in Multimedia Appendix 2. Sequence counts were rarefied to the smallest common value of 15 M using uniform sampling without replacement from the phyloseq package [38]. Sequence abundance was used as a proxy for taxonomic abundance. The α and β diversity were assessed using the inverse Simpson index [39] and Bray-Curtis dissimilarity [40], respectively. A *P* value lower than .05 was considered significant. We used a uniform manifold approximation and projection (UMAP) [41] technique in order to visually inspect potential clustering of the Bray-Curtis dissimilarity reduced to two dimensions. We estimated differential sequence abundance using a β-binomial model provided by the R package, corncob [42], with false discovery rate (<0.1)–adjusted *P* values. Clinical outcomes were assessed in a post hoc analysis by linear mixed-effect models [43]. Patients diagnosed with other diseases during the course of the study were excluded from the analyses if the disease affected their results; otherwise, they were included.

Results

Participants and Study Course

In total, 34 IBS patients were included in the study (Figure 2). A total of 3 (9%) patients dropped out before randomization, and an additional 9 (26%) IBS patients discontinued participation during the course of the year. Reasons for discontinuation were divorce or other stress-related events. In total, 5 (15%) IBS patients were diagnosed with other diseases during the study: 1 with microscopic colitis, 1 with neuroendocrine tumor, 2 with bile acid malabsorption, and 1 with breast cancer; 4 of these completed the study. A total of 6 HCs were included, all of whom completed the 1-year study. Baseline data of participants are shown in Table 1. The number of IBS patients enrolled, randomizations to interventions, and responders and nonresponders to interventions can be found in Figure 2. Out of 21 patients, 12 responded to LFD (57%) and 8 (38%) completed the reintroduction to high-FODMAP foods. Out of 21 patients, 13 (62%) responded to their first VSL#3 treatment and 7 (33%) responded to multiple VSL#3 treatments. A total of 180 fecal samples (156 from IBS patients and 24 from HCs) were analyzed. Out of 180 fecal samples, 7 (3.9%) were removed from the metagenomic data set due to medical treatments undergone by the participants (ie, antibiotics or cholestyramine treatment). An additional 22 samples (12.2%) were removed from the data set for analyses that excluded patients diagnosed with other diseases during the study.

Figure 2. Flowchart of inclusion, individualized treatment response, and 1-year follow-up. Response to a 4-week probiotic treatment or LFD was defined as a decrease in the IBS Severity Scoring System (IBS-SSS) of at least 50 points. *One patient did not manage to cross over to probiotic treatment; however, this patient completed 1 year on the web application. This participant is labelled as "discontinued" in the figure. **Diagnosed with another disease and treated medically in the outpatient clinic. FODMAPs: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS CC: IBS Constant Care.

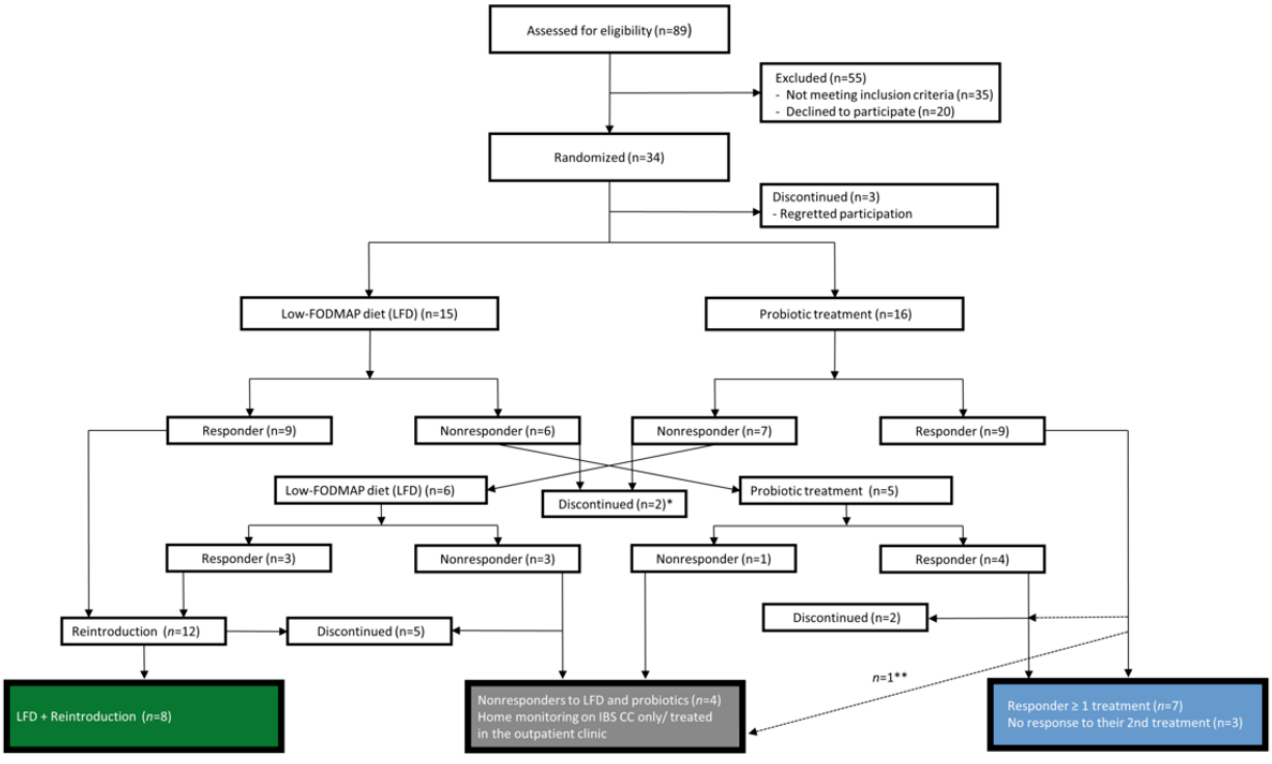


Table 1. Characteristics of participants at inclusion and upon randomization.

Characteristic	Inclusion (day 0; N=34 patients)	At randomization (week 4; n=31 patients)
Irritable bowel syndrome (IBS) patients		
IBS subtype, n (%)		
Mixed	15 (44)	15 (48)
Diarrhea	19 (56)	16 (52)
Gender, n (%)		
Female	23 (68)	22 (71)
Male	11 (32)	9 (29)
Smoking status, n (%)		
Never	18 (52)	18 (58)
Current	4 (12)	4 (13)
Previous	9 (26)	9 (29)
Missing	3 (9)	0 (0)
Disease course type, n (%)		
A (mild IBS with indolent course)	0 (0)	0 (0)
B (mild IBS with aggressive course)	11 (32)	11 (36)
C (chronic IBS with continuous course)	16 (47)	16 (52)
D (chronic IBS with intermittent course)	4 (12)	4 (13)
Missing	3 (9)	0 (0)
Age (years), median (IQR)	44.5 (27.5-51.5)	44 (26-50)
Patient-reported years with IBS, median (IQR)	5 (2.9-15.0)	5 (2.5-15.0)
BMI[a], median (IQR)	23.9 (21.9-26.4)	23.6 (21.2-26.1)
Fecal calprotectin (FC; mg/kg), median (IQR)	N/A[b]	53 (0-69)
IBS-SSS[c] score (0-500), median (IQR)	312 (243-370)	292 (225-356)
Bristol Stool Chart score (1-7), median (IQR)	5 (3-6)	5 (3-6)
Bowel movement frequency per day, median (IQR)	2 (1-5)	2 (2-4)
IBS-QoL[d] questionnaire score (0-100) , median (IQR)	54 (38-77)	53 (38-80)
Evaluation of IBS CC[e], including FC home test (1-10)[f], median (IQR)	N/A	8 (7-9)
Healthy controls (n=6)[g]		
Gender, n (%)		
Female	5 (83)	N/A
Male	1 (17)	N/A
Age (years), median (IQR)	46 (30.8-58.5)	N/A
BMI, median (IQR)	23.2 (20.8-25.1)	N/A
IBS-SSS score (0-500), median (IQR)	13.5 (0.75-22.25)	N/A
Bristol Stool Chart score (1-7), median (IQR)	4 (3.75-4.5)	N/A
Bowel movement frequency per day, median (IQR)	1.5 (1-2.25)	N/A
FC (mg/kg), median (IQR)	0 (0-72.75)	N/A
IBS-QoL questionnaire score (0-100), median (IQR)	100 (98.25-100)	N/A

[a]BMI is calculated as weight in kg divided by height in m^2.

[b]N/A: not applicable; data were not collected for this characteristic at this time point.

[c]IBS-SSS: IBS Severity Scoring System.

[d]IBS-QoL: IBS quality of life.

[e]IBS CC: IBS Constant Care.

[f]Satisfaction with the expanded IBS CC web application, including the FC home test, was evaluated by patients on a visual analog scale from 1 to 10, where 10 indicates the greatest possible satisfaction.

[g]Data for healthy controls were collected only at inclusion.

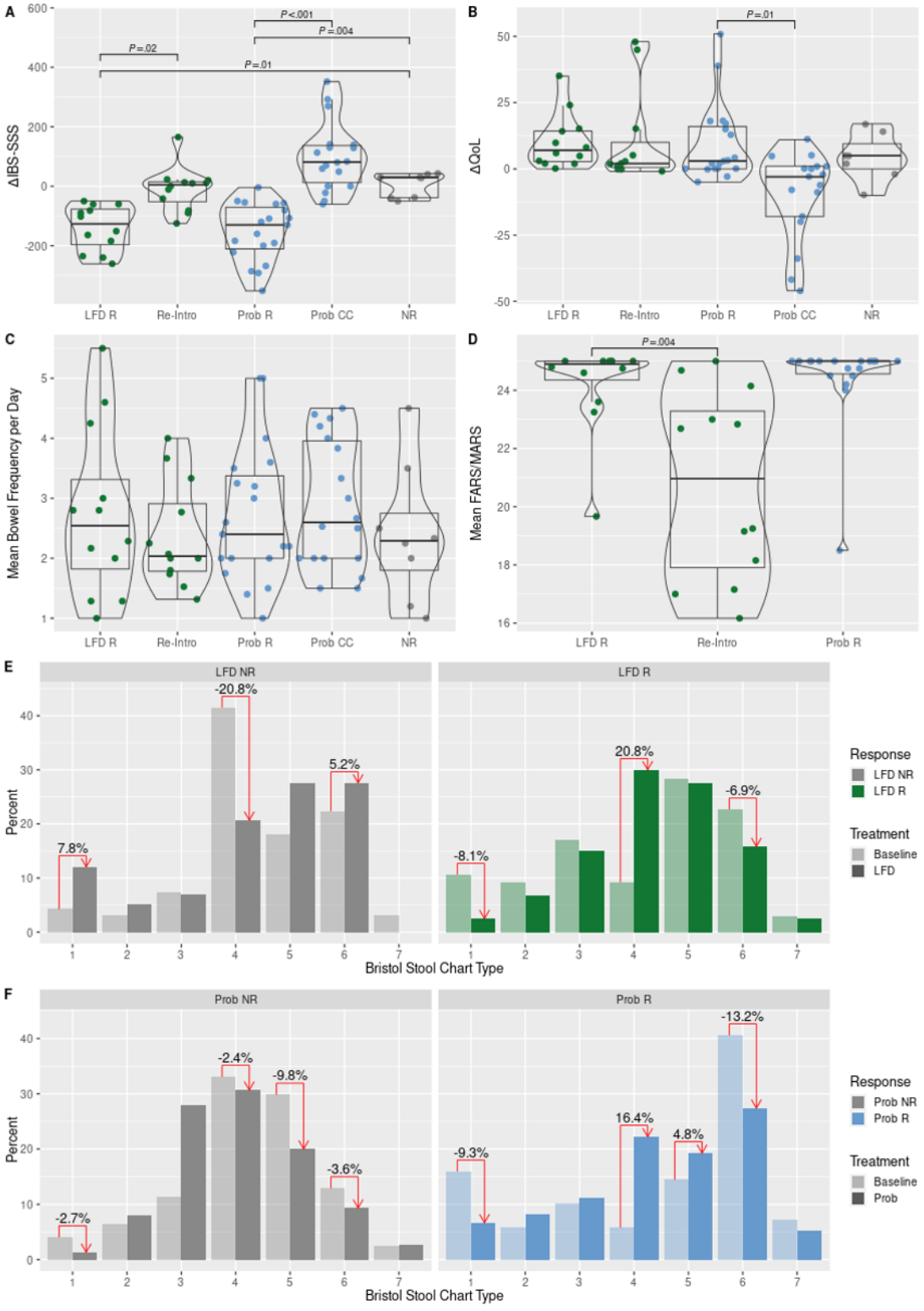
Treatment Duration and Effect Sizes of the LFD and Probiotic Treatments Based on Response Types

The median time spent dieting by LFD responders (12/21, 57%) was 36 days (IQR 28.00-41.25), and the median time spent for reintroduction of high-FODMAP foods (8/21, 38%) was 296 days (IQR 227.00-305.75). For "true" responders (ie, participants who responded to multiple probiotic treatments [7/21, 33%]; see Figure 2), the median number of probiotic treatments needed for maintaining symptom control during the course of 1 year was 3 (IQR 2.25-3.75), and the median number of days between probiotic treatments was 46.5 (IQR 26.25-65.75). Symptom remission (ie, an IBS-SSS lower than 175) was achieved by 8 out of 12 (67%) LFD responders and 5 out of 7 (71%) multiple probiotic responders, the latter corresponding to 12 out of 19 (63%) VSL#3 treatments bringing about symptom remission in probiotic responders.

As shown in Figure 3, A, significant decreases in IBS-SSS were observed for LFD and probiotic responders relative to nonresponders; both responder types were adherent to treatments (Figure 3, D). However, no significant difference in effect size between the two treatments was found; the median IBS-SSS effect sizes were –126.50 (IQR –196.75 to –76.75) for LFD responders and –130.00 (IQR –211.00 to –70.50) for probiotic responders (P>.99). The corresponding changes in QoL among LFD and probiotic responders was not significant compared to nonresponders, nor were the median increases in QoL significant between the two responder groups, which were 7 (IQR 2.75-14.20) for LFD responders and 3 (IQR 0-16.00) for probiotic responders (P>.99). During reintroduction to high-FODMAP foods, patients deviated significantly from the principles of the LFD (Figure 3, D; P=.004), resulting in a small but significant median increase in symptom severity relative to the time period where they were on the LFD (Figure 3, A). In median terms, probiotic responders experienced a significant increase in symptom severity between active probiotic treatments, which also resulted in a significant median decrease in QoL (Figure 3, A and B). No significant changes in bowel movements per day based on responder type were observed (Figure 3, C). LFD and probiotic responders showed a tendency of normalizing stool appearance to type 4 (Figure 3, E and F). These results are generally substantiated by fitting linear mixed-effect models of clinical metadata from the IBS CC database (ie, not considering delta values alone) to estimate effect sizes (eg, for responder groups). These post hoc models of QoL showed that LFD responders saw significant increases in their QoL while on the LFD (by an estimated 9.08), and even more during reintroduction (13.48) relative to their baseline QoL (an estimated 60.20). The model estimating the effect on severity scoring showed significant decreases for all responder types, including probiotic responders between active treatments. Outputs from the linear mixed-effect models are shown in Multimedia Appendix 2.

No significant changes in FC based on responder type were registered between baseline and 1-year follow-up (Multimedia Appendix 2). At baseline, none of the IBS patients reported an indolent disease course type (ie, type A). At 1-year follow-up, 42% (5/12) and 43% (3/7) of the LFD and probiotic responders, respectively, reported an indolent disease course type (ie, type A). None of the nonresponders reported an indolent course at 1-year follow-up.

Figure 3. Effect sizes for select patient-reported outcome measures based on responder types. LFD responders (LFD R, n=12); reintroduction (Re-Intro, n=12); probiotic responders (Prob R, n=7; "true" responders; 19 treatments); in between active probiotic treatments (Prob CC, n=7; 18 between periods); nonresponders (NR, n=4). A. Change in IBS-SSS. B. Change in IBS-QoL. C. Mean bowel movement frequency per day. D. Mean adherence for LFD responders (FARS) and during reintroduction (FARS) and for probiotic responders (MARS). E and F. Change in Bristol Stool Chart scores, expressed as percentages. FARS: FODMAP Adherence Report Scale; FODMAPs: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS: irritable bowel syndrome; IBS-SSS: IBS Severity Scoring System; LFD: low-FODMAP diet; MARS: Medical Adherence Report Scale; Prob: probiotic treatment; Prob CC: in between probiotic treatments, where patients are only measuring on IBS Constant Care (IBS CC; ie, not receiving any active treatments); QoL: quality of life; Re-Intro: resuming consumption of foods high in FODMAPs.



Reintroduction of High-FODMAP Foods

A total of 8 out of 12 (67%) participants completed the process of reintroducing high-FODMAP foods (Figure 2) and succeeded in developing an individualized diet at 1-year follow-up. They reintroduced a median of 14.50 (IQR 7.25-21.75) high-FODMAP foods, of which they categorized a median of 7 (IQR 3-13) foods as green (ie, not symptom-triggering foods), while they categorized 5 (IQR 4-8) foods as yellow and 2 (IQR 1-4) foods as red (ie, symptom-triggering foods). Common among symptom-triggering foods were wheat and rye bread, pasta, pointed cabbage, onion, garlic, leek, broccoli, green peas, cauliflower, kidney beans, chickpeas, sweet potatoes, avocado, mushrooms, apples, and apple juice.

Safety and Patient Satisfaction With IBS CC

No adverse events were registered for either intervention. The evaluation at 1-year follow-up showed overall satisfaction with the study, the support received during the study, and with IBS CC itself (further details can be found in Multimedia Appendix 2). During the 1-year study period, originating from both the personnel doing the daily web-based ward rounds and participants (n=33, including HCs), 887 web consultations and 781 text messages were registered.

Healthy Controls

HCs did not receive any intervention during the study period and were in IBS-SSS remission at inclusion and at 1-year follow-up. No significant change was detected (median 13.5 [IQR 19.25] vs 15.5 [IQR 14.75], *P*=.79). They had high IBS-QoL scores at inclusion and 1-year follow-up (median 100 [0.75] vs 100 [0.75], *P*>.99). The primary purpose for the inclusion of HCs was to generate microbiome profiles that could serve as a reference.

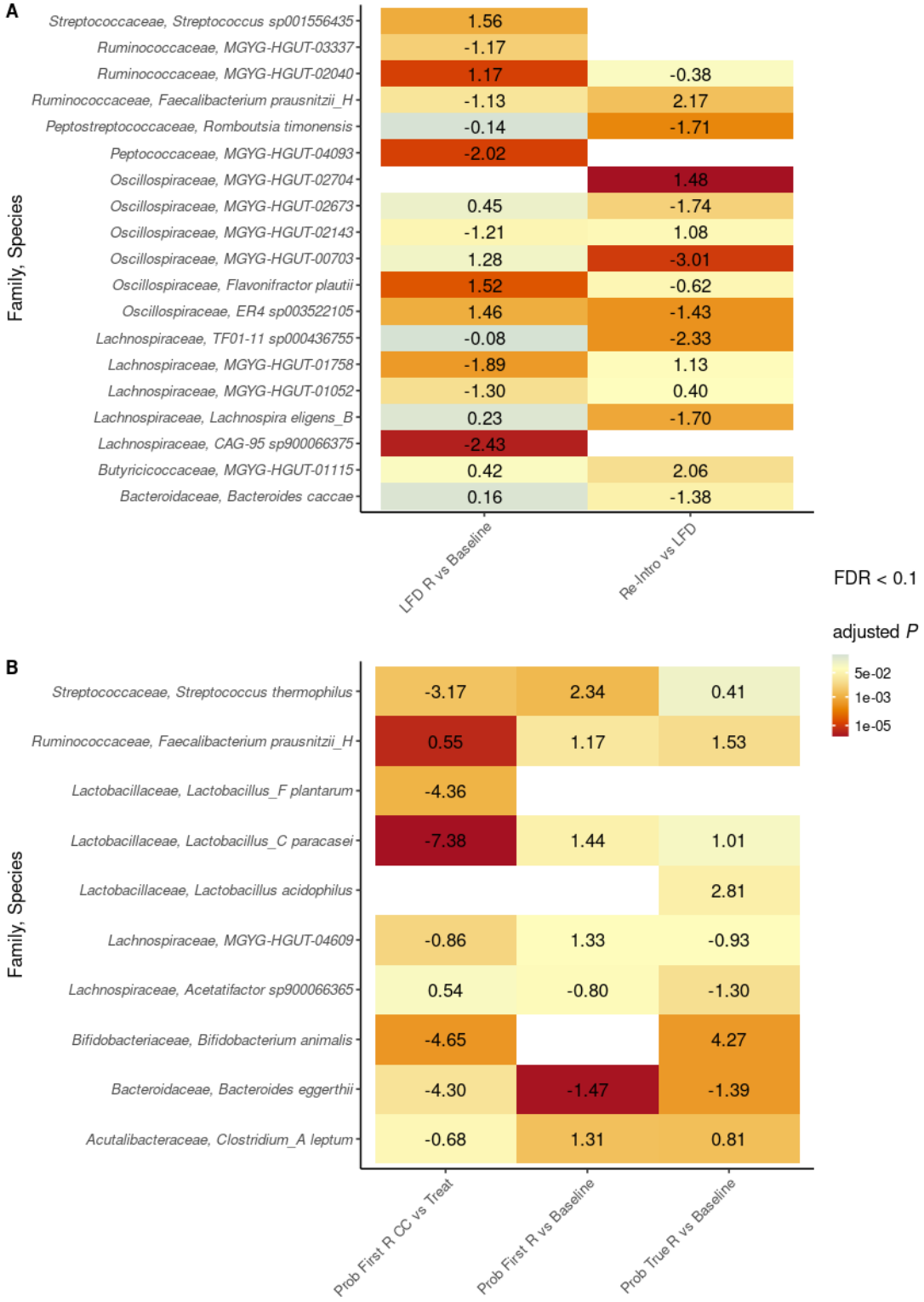
Basic Microbiota Descriptions in Relation to Diagnoses and Responder Types

A total of 10 million reads per sample were matched with high confidence to a median of 940 (IQR 18, range 821-1007) microbial species per sample. Median α diversity (measured by inverse Simpson index) was significantly greater for HCs (n=6) than for IBS patients (n=31), with median values of 51.60 (IQR 23.5) and 42.10 (IQR 16.1), respectively (*P*=.03; Multimedia Appendix 2). Across the entire study population, the Bray-Curtis dissimilarity within individuals was a median of 0.42 (IQR 0.14), and between individuals it was 0.72 (IQR 0.12; *P*<.001). There was also a significantly higher median of intraindividual Bray-Curtis dissimilarity in IBS-D compared to IBS-M patients (*P*=.03). UMAP plots based on IBS subtypes and responder types are shown in Multimedia Appendix 2. Those plots do not suggest any clusters, except for those given by the individuals themselves. The importance of individuals is also substantiated in the linear mixed-effect models estimating the change in α diversity for responder types (Multimedia Appendix 2). This model showed that approximately 24% of the variance was explained by the random effect of the individual and only 4% was explained by the fixed effect of responder type. In the model, nonresponders to the LFD had a lower inverse Simpson index that was close to significance (*P*=.07). No significant changes in α diversity for probiotic responders and nonresponders were found. Nonresponders to both probiotics and LFD showed a significantly higher median of intraindividual Bray-Curtis dissimilarity relative to probiotic responders (*P*=.01; Multimedia Appendix 2).

Differential Analyses of Species Abundances Based on Responder Types

No significant differential species abundances were found when comparing nonresponders to LFD and probiotic responders. Significant changes in relative species abundances induced by interventions (ie, LFD, reintroduction, and probiotic treatment) relative to baseline are shown in Figure 4 and further elaborated on in Multimedia Appendix 3. The corresponding Kyoto Encyclopedia of Genes and Genomes modules can be found in Multimedia Appendix 4.

Figure 4. Heat maps showing the 10 most differentially abundant species for low-FODMAP diet (LFD) comparisons (A) and the top five species for the probiotic (Prob) comparisons (B). A. LFD responders (LFD R, n=12) and reintroduction (Re-Intro, n=8). B. Probiotic responders. First-time responders (Prob First R) to treatment (Treat) and the time between active treatments (the period in between active probiotic treatment is abbreviated as CC; n=13). Second-time treatment responders ("true" responders, Prob True R; n=7; 19 treatments). The values shown in the heat maps denote the expected difference in the logit-transformed relative abundances between two samples from the groups being compared, controlling for the effect of individuals. The color scale represents the false discovery rate (FDR) (<0.1)–adjusted *P* values. The scale ranges from red to yellow to blue. Yellow marks the *P* value .05. All *P* values smaller than .05 tend toward red, and those larger than .05 tend toward blue. FODMAPs: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols.



Discussion

Principal Findings

In this study, we have shown that the new and expanded IBS CC web application for long-term management of IBS is practical, safe, and useful in clinical decision making. This was demonstrated by two year-long treatment interventions: the LFD (with subsequent reintroduction of high-FODMAP foods for LFD responders) and a probiotic treatment (repeated for 4 weeks every time a symptom flare-up occurred in probiotic responders). These therapeutics were found to be equally good at managing IBS symptoms in the short term and the long term. We also investigated participants' fecal microbiota, in order to try to predict response to the interventions, in the hope of being able to individualize future treatments for IBS; however, we were unable to make these predictions. Due to the small sample size of the study, larger ones are needed to confirm our results and would, ideally, include other medical options for managing IBS, especially for nonresponders to the LFD and probiotic treatments.

Sample Size and the Effect of the Individuals

Although the prevalence of IBS is high in westernized countries [1] such as Denmark, we were only able to include 34 out of an estimated target of 104 patients. This was mainly due to a high prevalence of comorbidities among eligible subjects that resulted in our excluding 64% of them. The high prevalence of comorbidities among eligible subjects is the topic of a newly published Danish nationwide study reporting that two-thirds of the Danish population, aged 16 years or older, suffer from one or more chronic conditions [44]. Our aim in excluding participants with multiple conditions, including other medical treatments, was to collect untarnished data for microbial analyses. The 5 patients with comorbidities or other diagnoses did not appear to affect the clinical data from IBS CC, nor did they affect the UMAP, inverse Simpson, or Bray-Curtis measures. However, these 5 patients did affect the results for the differential species abundance analyses based on responder types, possibly due to (1) the comorbidities themselves or (2) the fact that the individuals accounted for approximately 24% of the variance in the metagenomic data set. We aimed for including 104 IBS patients and 20 HCs. However, we did not manage to include our target population due to a high rate of co-morbidities among eligible subjects and implementation of new regulation (GDPR) in Denmark (data approval of the study took longer than expected).

Probiotic Treatment

The response rates, adherence, effect sizes of median reductions in symptoms, and increases in QoL for probiotic and LFD responders were similar (ie, not significantly different); however, probiotic responders experienced a significant increase in median symptom severity and a decrease in QoL in the times between the 4-week probiotic treatments, as compared to the effect obtained during active probiotic treatment. However, the linear mixed-effect model estimated that probiotic "true" responders sustained significant symptom control in between their 4-week treatments relative to their first measure (ie, the intercept). This is also partly explained by the study design, which instructed probiotic responders to contact the study investigator for a new 4-week probiotic treatment upon experiencing a symptom flare-up (ie, an increase in their IBS-SSS of at least 50 points). Sustaining IBS symptom control in between active probiotic treatments has, to our knowledge, not been investigated before among "true" responders. The concept of sustaining symptom control among "true" responders—with a median of three 4-week treatments across the year-long study—and the timing of new probiotic treatments using eHealth, require further investigation in larger, clinical web-based trials. The high dose (2 × 450 billion CFU per day) and short duration combination used in this study seems reasonable, based on a meta-analysis that concluded that (1) probiotics are effective and safe for IBS patients (moderate evidence) and (2) single probiotics at a higher daily dose and for a shorter duration (less than 8 weeks) seem to be more efficacious (needs further evidence) [45]. Although the higher dose and shorter duration were related to single-strain probiotics, we have also documented the effect of multi-strain VSL#3. Among the top five most differentially abundant bacterial species found when comparing samples from probiotic responders to their baseline samples, we identified four species (ie, *S thermophilus*, *L paracasei*, *L acidophilus*, and *B animalis*) that are representative of strains contained in the VSL#3 mixture. Furthermore, all four species were completely absent from baseline samples, which means that the consumption of VSL#3 is the most plausible explanation for their presence. Future long-term studies need to investigate whether it is safe for patients to stop probiotic treatment after 4 weeks and to only resume when symptom flare-ups occur. Abrupt suspension of probiotics have been hypothesized to increase host pathogen susceptibility by inducing gut dysbiosis [46].

LFD and Reintroduction of High-FODMAP Foods

Several key studies, meta-analyses, and reviews [10,47-50] have investigated the effects of an LFD on symptom reduction and changes in microbiota during an LFD [14,51-53]. Similar effect sizes and response rates to those we found have also been reported before [13,25,52]; however, the changes observed in the microbiota profile induced by the LFD in this study are not the same as reported in previous studies [14,51-53]. In one of these studies, easing of FODMAP restrictions (ie, reintroduction of foods high in FODMAPs) was recommended based on microbiota results indicating a significant decrease in total bacteria abundances, including significant changes in relative abundances of *Clostridium* cluster XIVa (reduced), *A muciniphila* (reduced), and *R torques* (increased) relative to a normal Australian diet [14]. Others have shown that the LFD, relative to sham diet advice, induces significant changes in the relative abundances of *Bifidobacterium* (reduced), an unclassified genus in the Ruminococcaceae family (reduced), and *Bacteroides* (increased) [53]. The same authors, in other papers, suggest that coadministration of probiotics with an LFD can restore *Bifidobacterium* [52], or that a β-Galactooligosaccharide supplement taken alongside the LFD can improve IBS symptoms [54]. Our results suggest that the LFD nonresponders tended to have reduced α diversity, as measured by the inverse Simpson index, whereas responders did not. *Faecalibacterium prausnitzii*, a commensal bacterium

of the Ruminococcaceae family that produces short-chain fatty acids (SCFAs) from dietary fiber [55], was significantly reduced by the LFD and restored by the reintroduction of foods high in FODMAPs. It should be noted in this context that it was *Faecalibacterium prausnitzii* H that was significantly increased; G, E, J, K, and I were decreased, while C and F were increased, but none significantly so (Multimedia Appendix 3). The main explanations for the discrepancies between studies investigating the effects of an LFD on microbiota include, but are not limited to, different DNA extraction protocols [56,57]; different storage conditions [58]; different parameters and bioinformatic analyses, including different comparisons "between groups" rather than changes "within groups"; and use of different sequencing methods [59].

Although significant median increases in symptoms were observed during reintroduction of high-FODMAP foods, in this study, and relative to the changes caused by the LFD, a post hoc analysis (ie, a linear mixed-effect model based on clinical metadata) suggests that sustained long-term symptom control and improvement of QoL could be obtained relative to patients' baseline. Sustained symptom control during reintroduction has recently been documented by other researchers in Italy, who concluded that the benefit of an LFD persisted during reintroduction (after 3 months) and at 6 months follow-up [60]. Based on this study's results, the combination of using a food preference approach for reintroduction and eHealth seems fit for long-term disease management. However, further eHealth studies with larger cohorts are needed to confirm these results.

IBS and Web-Based Management

Other researchers have suggested that web-based and app-based LFD and probiotic treatments may provide therapeutic benefits [61] and should be more widely implemented for IBS management [13]. This study supports these existing results and recommendations and adds further evidence of using eHealth for clinical decision making and long-term management.

However, it is our recommendation that future electronic-based IBS management include valid and quick point-of-care measures that can (1) predict response to treatments and (2) be correlated to symptom severity scores, both of which can support the patients and the personnel doing the daily web-based ward rounds in tailoring treatments and guiding their supervision. We were unable to predict response to treatments (ie, probiotic treatment and LFD) based on the microbiota alone. Others have also failed in this attempt in IBS using a different method (ie, 16S ribosomal RNA [rRNA] next-generation sequencing) than the one we used [53]. It could be that predicting a response to treatment and symptom activity would gain from larger cohorts and more detailed analyses of the gut ecosystem (eg, including other microbes, such as parasites [62]) and metabolites. A promising, low-cost, and noninvasive measure for predicting response to LFD and probiotic treatments in IBS patients is currently under development by researchers at King's College London that examines volatile organic compounds (VOCs) in feces (ie, VOC profiling) [63]; future evaluation of this method will hopefully determine whether it has clinical utility. A final, important aspect of research into long-term IBS management is the exclusion of other gastrointestinal diseases and comorbidities, as these frequently occur together [44]. A review by Kim et al [64] reported that FC had the highest sensitivity and specificity relative to IBD, whereas fecal SCFAs were most accurate relative to HCs. In the present study, neuroendocrine tumor showed itself with repeatedly elevated FC measures, but this measure did not indicate bile acid malabsorption or microscopic colitis. Future research might reveal new and valid point-of-care markers for detection of other diseases, for surveillance of IBS activity (ie, predictors for symptom relapse), and for predicting responses to therapeutics, thereby helping us tailor treatments and their timing, via improved surveillance, for patients.

Strengths and Limitations

The strength of this study is its use of an updated and expanded eHealth web application to monitor and treat IBS patients according to their treatment response across 1 year of follow-up. To our knowledge, this is the first eHealth trial to monitor the effect, prospectively and longitudinally, of multiple probiotic treatments and the LFD with subsequent reintroduction of foods high in FODMAPs, and to compare responses to patients' microbiota.

The main limitations of the study are its sample size and that it is based on a single study center. Another limitation is the use of the Rome III criteria rather than the Rome IV criteria for IBS [65]. Furthermore, gut microbiota were analyzed using fecal samples and metagenomic shotgun sequencing as a proxy. While metagenomic, whole-genome shotgun sequencing allows for a better genome resolution compared to 16S rRNA amplicon sequencing, organisms that are extremely GC-rich or -poor will be underrepresented [66]. Due to costs, samples were also sequenced at a lower depth (an average of 20 M reads), which limits the possibility of detecting microbes in very low abundances. In order to analyze the microbiota, we used a read-based mapping approach that is faster than metagenomic genome assembly, avoids the many problems that can occur during assembly, and provides full-genome information for low-abundance microbes that cannot be assembled otherwise [67]. However, this also meant that we could not capture the specific capabilities of each microbe in a patient, encoded by their particular genome. The *in situ* genome sequences likely differ compared to our reference due to, for example, horizontal gene transfer.

Conclusions

In conclusion, we have shown the IBS CC web application to be practical, safe, and useful for clinical decision making in the long-term management of IBS. A probiotic treatment, VSL#3, and the LFD were equally efficacious in short- and long-term treatment strategies for managing IBS. We failed to predict response to treatments based on patients' fecal microbiota, which might someday help individualize treatments. To confirm our results, larger studies are needed and would ideally combine LFD and probiotic treatments or other medical options for managing IBS, especially for nonresponders to the LFD and probiotic treatments.

Acknowledgments

We wish to thank Thomas Janum (J-Consult) for support in developing the IBS CC application, integration of CalproSmart, and the export function from the IBC CC database. Grants supporting this study were provided by North Zealand University Hospital, Ferring Pharmaceuticals, the Handelsgartner Ove William Buhl Olesen and ægtefælle fru Edith Buhl Olesen memorial fund, the Dagmar and Ivar Enevoldsen memorial fund, and the Crohn's & Colitis Patient Society, Denmark. This study was also supported, in part, by Unseen Bio ApS and Calpro AS.

Authors' Contributions

DVA prepared the manuscript, which was critically reviewed by all coauthors. DVA, JB, and PM designed the study. DVA, PM, MB, AH, EF, SS, NDV, TR, and SF conducted the study. DVA, MEB, and PM had full access to the data in the study and take full responsibility for their integrity. DVA prepared the data tables and MEB performed the statistical analyses. All authors approved the final version of the manuscript.

Conflicts of Interest

MEB and CL are cofounders of Unseen Bio ApS, a company that offers microbiome profiling to consumers. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1253 KB - jmir_v23i12e30291_app1.pdf]

Multimedia Appendix 2
Supplementary elaborations, tables, and figures.
[PDF File (Adobe PDF File), 2518 KB - jmir_v23i12e30291_app2.pdf]

Multimedia Appendix 3
Differential species abundances according to responder types.
[XLSX File (Microsoft Excel File), 1244 KB - jmir_v23i12e30291_app3.xlsx]

Multimedia Appendix 4
Differential Kyoto Encyclopedia of Genes and Genomes modules according to responder types.
[XLSX File (Microsoft Excel File), 196 KB - jmir_v23i12e30291_app4.xlsx]

References

1. Canavan C, West J, Card T. The epidemiology of irritable bowel syndrome. Clin Epidemiol 2014;6:71-80 [FREE Full text] [doi: 10.2147/CLEP.S40245] [Medline: 24523597]
2. Corsetti M, Whorwell P. The global impact of IBS: Time to think about IBS-specific models of care? Therap Adv Gastroenterol 2017 Sep;10(9):727-736 [FREE Full text] [doi: 10.1177/1756283X17718677] [Medline: 28932273]
3. Simrén M, Barbara G, Flint HJ, Spiegel BMR, Spiller RC, Vanner S, Rome Foundation Committee. Intestinal microbiota in functional bowel disorders: A Rome Foundation report. Gut 2013 Jan;62(1):159-176 [FREE Full text] [doi: 10.1136/gutjnl-2012-302167] [Medline: 22730468]
4. Mayer EA. The role of gut-brain interactions in influencing symptoms of irritable bowel syndrome. Gastroenterol Hepatol (N Y) 2018 Jan;14(1):44-46 [FREE Full text] [Medline: 29491761]
5. Thabane M, Marshall JK. Post-infectious irritable bowel syndrome. World J Gastroenterol 2009 Aug 07;15(29):3591-3596 [FREE Full text] [doi: 10.3748/wjg.15.3591] [Medline: 19653335]
6. Jones MP, Chey WD, Singh S, Gong H, Shringarpure R, Hoe N, et al. A biomarker panel and psychological morbidity differentiates the irritable bowel syndrome from health and provides novel pathophysiological leads. Aliment Pharmacol Ther 2014 Feb;39(4):426-437 [FREE Full text] [doi: 10.1111/apt.12608] [Medline: 24387672]
7. Ankersen DV, Weimers P, Burisch J. Whats 'app-ening': The help of new technologies in nutrition in digestive diseases. Curr Opin Clin Nutr Metab Care 2017 Sep;20(5):426-431. [doi: 10.1097/MCO.0000000000000399] [Medline: 28768297]
8. Pedersen N. EHealth: Self-management in inflammatory bowel disease and in irritable bowel syndrome using novel constant-care web applications. EHealth by constant-care in IBD and IBS. Dan Med J 2015 Dec;62(12):B5168 [FREE Full text] [Medline: 26621403]

9. Whelan K, Martin LD, Staudacher HM, Lomer MCE. The low FODMAP diet in the management of irritable bowel syndrome: An evidence-based review of FODMAP restriction, reintroduction and personalisation in clinical practice. J Hum Nutr Diet 2018 Apr;31(2):239-255. [doi: 10.1111/jhn.12530] [Medline: 29336079]
10. Altobelli E, Del Negro V, Angeletti P, Latella G. Low-FODMAP diet improves irritable bowel syndrome symptoms: A meta-analysis. Nutrients 2017 Aug 26;9(9):940 [FREE Full text] [doi: 10.3390/nu9090940] [Medline: 28846594]
11. Maagaard L, Ankersen DV, Végh Z, Burisch J, Jensen L, Pedersen N, et al. Follow-up of patients with functional bowel symptoms treated with a low FODMAP diet. World J Gastroenterol 2016 Apr 21;22(15):4009-4019 [FREE Full text] [doi: 10.3748/wjg.v22.i15.4009] [Medline: 27099444]
12. Weynants A, Goossens L, Genetello M, De Looze D, Van Winckel M. The long-term effect and adherence of a low fermentable oligosaccharides disaccharides monosaccharides and polyols (FODMAP) diet in patients with irritable bowel syndrome. J Hum Nutr Diet 2020 Apr;33(2):159-169. [doi: 10.1111/jhn.12706] [Medline: 31637777]
13. Pedersen N, Andersen NN, Végh Z, Jensen L, Ankersen DV, Felding M, et al. Ehealth: Low FODMAP diet vs Lactobacillus rhamnosus GG in irritable bowel syndrome. World J Gastroenterol 2014 Nov 21;20(43):16215-16226 [FREE Full text] [doi: 10.3748/wjg.v20.i43.16215] [Medline: 25473176]
14. Halmos EP, Christophersen CT, Bird AR, Shepherd SJ, Gibson PR, Muir JG. Diets that differ in their FODMAP content alter the colonic luminal microenvironment. Gut 2015 Jan;64(1):93-100. [doi: 10.1136/gutjnl-2014-307264] [Medline: 25016597]
15. Staudacher HM, Irving PM, Lomer MCE, Whelan K. Mechanisms and efficacy of dietary FODMAP restriction in IBS. Nat Rev Gastroenterol Hepatol 2014 Apr;11(4):256-266. [doi: 10.1038/nrgastro.2013.259] [Medline: 24445613]
16. Pittayanon R, Lau JT, Yuan Y, Leontiadis GI, Tse F, Surette M, et al. Gut microbiota in patients with irritable bowel syndrome-A systematic review. Gastroenterology 2019 Jul;157(1):97-108. [doi: 10.1053/j.gastro.2019.03.049] [Medline: 30940523]
17. Probiotics Market Size, Share & Trends Analysis Report by Product (Food & Beverages, Dietary Supplements), by Ingredient (Bacteria, Yeast), by End Use, by Distribution Channel, and Segment Forecasts, 2019 - 2025. San Francisco, CA: Grand View Research; 2019. URL: https://www.grandviewresearch.com/industry-analysis/probiotics-market [accessed 2021-04-06]
18. Yang M, Yu Y, Lei P, Yuan J. Comparative efficacy and safety of probiotics for the treatment of irritable bowel syndrome: A systematic review and network meta-analysis protocol. BMJ Open 2019 Dec 02;9(12):e027376 [FREE Full text] [doi: 10.1136/bmjopen-2018-027376] [Medline: 31796470]
19. Hookway C, Buckner S, Crosland P, Longson D. Irritable bowel syndrome in adults in primary care: Summary of updated NICE guidance. BMJ 2015 Feb 25;350:h701. [doi: 10.1136/bmj.h701] [Medline: 25716701]
20. McFarland LV, Dublin S. Meta-analysis of probiotics for the treatment of irritable bowel syndrome. World J Gastroenterol 2008 May 07;14(17):2650-2661 [FREE Full text] [doi: 10.3748/wjg.14.2650] [Medline: 18461650]
21. Didari T, Mozaffari S, Nikfar S, Abdollahi M. Effectiveness of probiotics in irritable bowel syndrome: Updated systematic review with meta-analysis. World J Gastroenterol 2015 Mar 14;21(10):3072-3084 [FREE Full text] [doi: 10.3748/wjg.v21.i10.3072] [Medline: 25780308]
22. Wieërs G, Belkhir L, Enaud R, Leclercq S, Philippart de Foy J, Dequenne I, et al. How probiotics affect the microbiota. Front Cell Infect Microbiol 2019;9:454 [FREE Full text] [doi: 10.3389/fcimb.2019.00454] [Medline: 32010640]
23. IBS Constant Care. URL: https://ibs.constant-care.com/ [accessed 2021-11-29]
24. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: A simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 Apr;11(2):395-402 [FREE Full text] [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]
25. Pedersen N, Vegh Z, Burisch J, Jensen L, Ankersen DV, Felding M, et al. Ehealth monitoring in irritable bowel syndrome patients treated with low fermentable oligo-, di-, mono-saccharides and polyols diet. World J Gastroenterol 2014 Jun 07;20(21):6680-6684 [FREE Full text] [doi: 10.3748/wjg.v20.i21.6680] [Medline: 24914395]
26. Carlsen K, Jakobsen C, Houen G, Kallemose T, Paerregaard A, Riis LB, et al. Self-managed eHealth disease monitoring in children and adolescents with inflammatory bowel disease: A randomized controlled trial. Inflamm Bowel Dis 2017 Mar;23(3):357-365. [doi: 10.1097/MIB.0000000000001026] [Medline: 28221247]
27. Pedersen N, Thielsen P, Martinsen L, Bennedsen M, Haaber A, Langholz E, et al. eHealth: Individualization of mesalazine treatment through a self-managed web-based solution in mild-to-moderate ulcerative colitis. Inflamm Bowel Dis 2014 Dec;20(12):2276-2285. [doi: 10.1097/MIB.0000000000000199] [Medline: 25248002]
28. Patrick D, Drossman D, Frederick I, DiCesare J, Puder K. Quality of life in persons with irritable bowel syndrome: Development and validation of a new measure. Dig Dis Sci 1998 Feb;43(2):400-411. [doi: 10.1023/a:1018831127942] [Medline: 9512138]
29. bcl2fastq2 Conversion Software v2.20. Illumina. 2017. URL: https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html [accessed 2021-04-01]
30. Jiang H, Lei R, Ding S, Zhu S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 2014 Jun 12;15:182 [FREE Full text] [doi: 10.1186/1471-2105-15-182] [Medline: 24925680]
31. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol 2017 Apr 11;35(4):316-319. [doi: 10.1038/nbt.3820] [Medline: 28398311]

32. Andrews S. FastQC. Babraham Bioinformatics. URL: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 2021-04-01]
33. Chen S, Zhou Y, Chen Y, Gu J. fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018 Sep 01;34(17):i884-i890 [FREE Full text] [doi: 10.1093/bioinformatics/bty560] [Medline: 30423086]
34. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019 Nov 28;20(1):257. [doi: 10.1186/s13059-019-1891-0] [Medline: 31779668]
35. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016 Oct 01;32(19):3047-3048 [FREE Full text] [doi: 10.1093/bioinformatics/btw354] [Medline: 27312411]
36. Mitchell A, Almeida A, Beracochea M, Boland M, Burgin J, Cochrane G, et al. MGnify: The microbiome analysis resource in 2020. Nucleic Acids Res 2020 Jan 08;48(D1):D570-D578 [FREE Full text] [doi: 10.1093/nar/gkz1035] [Medline: 31696235]
37. Lu J, Breitwieser F, Thielen P, Salzberg S. Bracken: Estimating species abundance in metagenomics data. PeerJ Comput Sci 2017:e104 [FREE Full text] [doi: 10.7717/peerj-cs.104]
38. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8(4):e61217 [FREE Full text] [doi: 10.1371/journal.pone.0061217] [Medline: 23630581]
39. Smith B, Wilson JB. A consumer's guide to evenness indices. Oikos 1996 May;76(1):70-82. [doi: 10.2307/3545749]
40. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci 2003 Apr 09;14(6):927-930. [doi: 10.1111/j.1654-1103.2003.tb02228.x]
41. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 2018 Sep;3(29):861. [doi: 10.21105/joss.00861]
42. Martin BD, Witten D, Willis AD. Modeling microbial abundances and dysbiosis with beta-binomial regression. Ann Appl Stat 2020 Mar;14(1):94-115 [FREE Full text] [doi: 10.1214/19-aoas1283] [Medline: 32983313]
43. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
44. Hvidberg MF, Johnsen SP, Davidsen M, Ehlers L. A nationwide study of prevalence rates and characteristics of 199 chronic conditions in Denmark. Pharmacoecon Open 2020 Jun;4(2):361-380 [FREE Full text] [doi: 10.1007/s41669-019-0167-7] [Medline: 31342402]
45. Li B, Liang L, Deng H, Guo J, Shu H, Zhang L. Efficacy and safety of probiotics in irritable bowel syndrome: A systematic review and meta-analysis. Front Pharmacol 2020;11:332 [FREE Full text] [doi: 10.3389/fphar.2020.00332] [Medline: 32317962]
46. Liu Z, Liu W, Ran C, Hu J, Zhou Z. Abrupt suspension of probiotics administration may increase host pathogen susceptibility by inducing gut dysbiosis. Sci Rep 2016 Mar 17;6:23214 [FREE Full text] [doi: 10.1038/srep23214] [Medline: 26983596]
47. Halmos EP, Power VA, Shepherd SJ, Gibson PR, Muir JG. A diet low in FODMAPs reduces symptoms of irritable bowel syndrome. Gastroenterology 2014 Jan;146(1):67-75.e5. [doi: 10.1053/j.gastro.2013.09.046] [Medline: 24076059]
48. Staudacher H, Whelan K, Irving P, Lomer M. Comparison of symptom response following advice for a diet low in fermentable carbohydrates (FODMAPs) versus standard dietary advice in patients with irritable bowel syndrome. J Hum Nutr Diet 2011 Oct;24(5):487-495. [doi: 10.1111/j.1365-277X.2011.01162.x] [Medline: 21615553]
49. Gearry R, Skidmore P, O'Brien L, Wilkinson T, Nanayakkara W. Efficacy of the low FODMAP diet for treating irritable bowel syndrome: The evidence to date. Clin Exp Gastroenterol 2016 Jun:131-142 [FREE Full text] [doi: 10.2147/ceg.s86798]
50. van Lanen A, de Bree A, Greyling A. Efficacy of a low-FODMAP diet in adult irritable bowel syndrome: A systematic review and meta-analysis. Eur J Nutr 2021 Sep;60(6):3505-3522 [FREE Full text] [doi: 10.1007/s00394-020-02473-0] [Medline: 33585949]
51. Staudacher H, Lomer M, Anderson J, Barrett JS, Muir JG, Irving PM, et al. Fermentable carbohydrate restriction reduces luminal bifidobacteria and gastrointestinal symptoms in patients with irritable bowel syndrome. J Nutr 2012 Aug;142(8):1510-1518. [doi: 10.3945/jn.112.159285] [Medline: 22739368]
52. Staudacher HM, Lomer MC, Farquharson FM, Louis P, Fava F, Franciosi E, et al. A diet low in FODMAPs reduces symptoms in patients with irritable bowel syndrome and a probiotic restores Bifidobacterium species: A randomized controlled trial. Gastroenterology 2017 Oct;153(4):936-947. [doi: 10.1053/j.gastro.2017.06.010] [Medline: 28625832]
53. Staudacher HM, Scholz M, Lomer MC, Ralph FS, Irving PM, Lindsay JO, et al. Gut microbiota associations with diet in irritable bowel syndrome and the effect of low FODMAP diet and probiotics. Clin Nutr 2021 Apr;40(4):1861-1870. [doi: 10.1016/j.clnu.2020.10.013] [Medline: 33183883]
54. Wilson B, Rossi M, Kanno T, Parkes GC, Anderson S, Mason AJ, et al. β-Galactooligosaccharide in conjunction with low FODMAP diet improves irritable bowel syndrome symptoms but reduces fecal Bifidobacteria. Am J Gastroenterol 2020 Jun;115(6):906-915. [doi: 10.14309/ajg.0000000000000641] [Medline: 32433273]
55. Lopez-Siles M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Faecalibacterium prausnitzii: From microbiology to diagnostics and prognostics. ISME J 2017 Apr;11(4):841-852 [FREE Full text] [doi: 10.1038/ismej.2016.176] [Medline: 28045459]

56. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017 Nov;35(11):1069-1076. [doi: 10.1038/nbt.3960] [Medline: 28967887]
57. Rintala A, Pietilä S, Munukka E, Eerola E, Pursiheimo J, Laiho A, et al. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. J Biomol Tech 2017 Apr;28(1):19-30 [FREE Full text] [doi: 10.7171/jbt.17-2801-003] [Medline: 28260999]
58. Gorzelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M, Gibson DL. Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. PLoS One 2015;10(8):e0134802 [FREE Full text] [doi: 10.1371/journal.pone.0134802] [Medline: 26252519]
59. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, et al. Critical Assessment of Metagenome Interpretation-A benchmark of metagenomics software. Nat Methods 2017 Nov;14(11):1063-1071 [FREE Full text] [doi: 10.1038/nmeth.4458] [Medline: 28967888]
60. Gravina AG, Dallio M, Romeo M, Di Somma A, Cotticelli G, Loguercio C, et al. Adherence and effects derived from FODMAP diet on irritable bowel syndrome: A real life evaluation of a large follow-up observation. Nutrients 2020 Mar 27;12(4):928 [FREE Full text] [doi: 10.3390/nu12040928] [Medline: 32230832]
61. Rafferty AJ, Hall R, Johnston CS. A novel mobile app (Heali) for disease treatment in participants with irritable bowel syndrome: Randomized controlled pilot trial. J Med Internet Res 2021 Mar 02;23(3):e24134 [FREE Full text] [doi: 10.2196/24134] [Medline: 33650977]
62. Krogsgaard L, Andersen L, Johannesen T, Engsbro AL, Stensvold CR, Nielsen HV, et al. Characteristics of the bacterial microbiome in association with common intestinal parasites in irritable bowel syndrome. Clin Transl Gastroenterol 2018 Jun 19;9(6):161 [FREE Full text] [doi: 10.1038/s41424-018-0027-2] [Medline: 29915224]
63. Rossi M, Aggio R, Staudacher HM, Lomer MC, Lindsay JO, Irving P, et al. Volatile organic compounds in feces associate with response to dietary intervention in patients with irritable bowel syndrome. Clin Gastroenterol Hepatol 2018 Mar;16(3):385-391.e1. [doi: 10.1016/j.cgh.2017.09.055] [Medline: 28993261]
64. Kim JH, Lin E, Pimentel M. Biomarkers of irritable bowel syndrome. J Neurogastroenterol Motil 2017 Jan 30;23(1):20-26 [FREE Full text] [doi: 10.5056/jnm16135] [Medline: 27817184]
65. Schmulson MJ, Drossman DA. What is new in Rome IV. J Neurogastroenterol Motil 2017 Apr 30;23(2):151-163 [FREE Full text] [doi: 10.5056/jnm16214] [Medline: 28274109]
66. Browne P, Nielsen T, Kot W, Aggerholm A, Gilbert MTP, Puetz L, et al. GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. Gigascience 2020 Feb 01;9(2):giaa008 [FREE Full text] [doi: 10.1093/gigascience/giaa008] [Medline: 32052832]
67. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017 Sep 12;35(9):833-844. [doi: 10.1038/nbt.3935] [Medline: 28898207]

Abbreviations

BSC: Bristol Stool Chart
CE: Conformité Européenne
CFU: colony-forming units
FARS: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP) Adherence Report Scale
FC: fecal calprotectin
FODMAP: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols
HC: healthy control
IBD: inflammatory bowel disease
IBS: irritable bowel syndrome
IBS CC: irritable bowel syndrome-constant care
IBS-D: irritable bowel syndrome–diarrhea
IBS-M: irritable bowel syndrome–mixed typed
IBS-QoL: irritable bowel syndrome quality of life
IBS-SSS: Irritable Bowel Syndrome Severity Scoring System
LFD: low-fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP) diet
MARS: Medical Adherence Report Scale
NICE: National Institute for Health and Care Excellence
PROM: patient-reported outcome measure
QoL: quality of life
rRNA: ribosomal RNA
SCFA: short-chain fatty acid
subsp: subspecies

UMAP: uniform manifold approximation and projection
VAS: visual analog scale
VOC: volatile organic compound

Edited by R Kukafka, G Eysenbach; submitted 09.05.21; peer-reviewed by S Weinland, J Knitza; comments to author 03.06.21; revised version received 22.06.21; accepted 25.10.21; published 14.12.21.

Please cite as:
Ankersen DV, Weimers P, Bennedsen M, Haaber AB, Fjordside EL, Beber ME, Lieven C, Saboori S, Vad N, Rannem T, Marker D, Paridaens K, Frahm S, Jensen L, Rosager Hansen M, Burisch J, Munkholm P
Long-Term Effects of a Web-Based Low-FODMAP Diet Versus Probiotic Treatment for Irritable Bowel Syndrome, Including Shotgun Analyses of Microbiota: Randomized, Double-Crossover Clinical Trial
J Med Internet Res 2021;23(12):e30291
URL: https://www.jmir.org/2021/12/e30291
doi:10.2196/30291
PMID:34904950

Original Paper

Web-Based Return of Individual Patient-Reported Outcome Results Among Patients With Lymphoma: Randomized Controlled Trial

Simone Oerlemans[1*], PhD; Lindy Paulina Johanna Arts[1*], PhD; Jacobien M Kieffer[2], PhD; Judith Prins[3], PhD; Mels Hoogendoorn[4], MD, PhD; Marjolein van der Poel[5], MD, PhD; Ad Koster[6], MD; Chantal Lensen[7], MD; Wendy Bernadina Catharina Stevens[8], MD, PhD; Djamila Issa[9], MD; Johannes F M Pruijt[9], MD, PhD; Margriet Oosterveld[10], MD, PhD; René van der Griend[11], MD, PhD; Marten Nijziel[12], MD, PhD; Lidwine Tick[13], MD, PhD; Eduardus F M Posthuma[14,15], MD, PhD; Lonneke V van de Poll-Franse[1,2,16], PhD

[1]Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Utrecht, Netherlands

[2]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, Netherlands

[3]Department of Medical Psychology, Radboud University Medical Center, Nijmegen, Netherlands

[4]Department of Hematology, Medical Center Leeuwarden, Leeuwarden, Netherlands

[5]Department of Internal Medicine, Division of Hematology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, Netherlands

[6]Department of Internal Medicine, VieCuri Medical Centre, Venlo/Venray, Netherlands

[7]Department of Internal Medicine, Bernhoven Hospital, Uden, Netherlands

[8]Department of Hematology, Radboud University Medical Center, Nijmegen, Netherlands

[9]Department of Internal Medicine, Jeroen Bosch Hospital, 's-Hertogenbosch, Netherlands

[10]Department of Internal Medicine, Canisius-Wilhelmina Hospital, Nijmegen, Netherlands

[11]Department of Internal Medicine, Diakonessenhuis, Utrecht/Zeist, Netherlands

[12]Department of Internal Medicine, Catharina Hospital, Eindhoven, Netherlands

[13]Department of Internal Medicine, Máxima Medical Centre, Veldhoven, Netherlands

[14]Department of Internal Medicine, Reinier de Graaf Group, Delft, Netherlands

[15]Department of Internal Medicine, Leiden University Medical Center, Leiden, Netherlands

[16]Center of Research on Psychological and Somatic disorders, Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

[*]these authors contributed equally

Corresponding Author:
Simone Oerlemans, PhD
Department of Research and Development
Netherlands Comprehensive Cancer Organisation
Godebaldkwartier 419, 3511 DT
Utrecht
Netherlands
Phone: 31 088 234 60 00
Email: s.oerlemans@iknl.nl

Abstract

Background: There has been a cultural shift toward patient engagement in health, with a growing demand from patients to access their results.

Objective: The Lymphoma Intervention (LIVE) trial is conducted to examine the impact of return of individual patient-reported outcome (PRO) results and a web-based self-management intervention on psychological distress, self-management, satisfaction with information, and health care use in a population-based setting.

Methods: Return of PRO results included comparison with age- and sex-matched peers and was built into the Patient-Reported Outcomes Following Initial Treatment and Long-Term Evaluation of Survivorship registry. The self-management intervention is an adaptation of a fully automated evidence-based intervention for breast cancer survivors. Patients with lymphoma who

completed the web-based questionnaire were equally randomized to care as usual, return of PRO results, and return of PRO results plus self-management intervention. Patients completed questionnaires 9 to 18 months after diagnosis (T0; n=227), 4 months (T1; n=190), 12 months (T2; n=170), and 24 months (T3; n=98).

Results: Of all invited patients, 51.1% (456/892) responded and web-based participants (n=227) were randomly assigned to care as usual (n=76), return of PRO results (n=74), or return of PRO results and access to Living with lymphoma (n=77). Return of PRO results was viewed by 76.7% (115/150) of those with access. No statistically significant differences were observed for psychological distress, self-management, satisfaction with information provision, and health care use between patients who received PRO results and those who did not (*P*>.05). Use of the self-management intervention was low (2/76, 3%), and an effect could therefore not be determined.

Conclusions: Return of individual PRO results seems to meet patients' wishes but had no beneficial effects on patient outcome. No negative effects were found when individual PRO results were disclosed, and the return of individual PRO results can therefore be safely implemented in daily clinical practice.

Trial Registration: Netherlands Trial Register NTR5953; https://www.trialregister.nl/trial/5790

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-017-1943-2

(J Med Internet Res 2021;23(12):e27886) doi:10.2196/27886

KEYWORDS

lymphoma; patient-reported outcomes; return of individual results; randomized controlled trial; self-management

Introduction

Background

Patients with lymphoma are at risk of experiencing adverse effects of cancer and its treatment, such as fatigue, cognitive problems, and neuropathy [1-4]. In addition, up to a quarter experience persistent levels of anxiety, depressive symptoms, and fear [5], also known as psychological distress. Both symptoms and psychological distress may be exacerbated when information and supportive care are unavailable [6,7]. This may, in turn, lead to increased health care use [8]. Regular screening of physical and psychosocial symptoms using patient-reported outcomes (PROs) could increase awareness and recognition of symptoms and can contribute to their management [9-13].

Since the past decade, there has been a cultural shift toward patient engagement in health, with a growing demand from patients to access their individual results [14-17]. Returning individual PRO results enables patients to monitor their functioning and create awareness of symptoms. Furthermore, it offers patients the opportunity to compare their scores with peers [15] to evaluate if their scores are *normal* and to incorporate this information into personal decision-making [16]. However, some clinicians have expressed reservations about disclosing PRO research results to patients, as it may cause patients to become more concerned and lead to increased health care use by patients and higher workload for clinicians. Therefore, it is important to investigate whether patients will be in a better or worse condition with their PRO results being disclosed and which patients wish to receive their PRO results.

To improve health outcomes, subsequent steps such as self-management interventions are expected to be necessary in addition to monitoring and returning PRO results [18]. Self-management interventions intend to enhance patients' knowledge and skills and empower them to play an active role in the management of their disease and its consequences [19,20]. Studies on the outcomes of self-management are mainly based on patients with solid cancers and are not consistent, with some randomized controlled trials (RCTs) showing no effects [21-23]. In addition, most self-management interventions are not specifically aimed at patients with lymphoma and are found to be effective among selected groups of patients. Little is known about the effects of such interventions in a population-based setting within an unselected group of patients.

Objectives

To investigate the effect of (1) return of individual PRO research results, including comparison with peers [24], and (2) a web-based self-management intervention Living with lymphoma, the Lymphoma Intervention (LIVE) trial was performed. The primary objective of the LIVE RCT was to examine the effects of return of PRO results to patients with or without access to Living with lymphoma on self-management, satisfaction with information, and psychological distress in a population-based setting of patients with lymphoma [24]. We hypothesized that those with access to the return of PRO results or access to LIVE would have higher levels of self-management and satisfaction with information and lower levels of psychological distress. On the basis of new insights that psychologically distressed patients reported increased health care use [8], we also investigated the effects of return of PRO results to patients with or without access to Living with lymphoma on health care use (secondary objective). Furthermore, we explored sociodemographic, clinical, and psychological differences between patients who viewed their individual PRO results and those who did not.

Methods

Design and Participants

This RCT was embedded in the population-based Patient-Reported Outcomes Following Initial Treatment and Long-Term Evaluation of Survivorship (PROFILES) registry [25]. PROFILES enables PRO data collection management and links PRO data to clinical data from the Netherlands Cancer Registry (NCR).

Between October 2015 and February 2019, patients diagnosed with lymphoma (Hodgkin lymphoma, non-Hodgkin lymphoma, or chronic lymphocytic leukemia), as defined by the International Classification of Diseases for Oncology-3 codes [26], from 13 hospitals in the Netherlands were selected for participation 9 to 18 months after diagnosis. The NCR registers all newly diagnosed patients with cancer in the Netherlands within the first year after diagnosis and routinely collects detailed data on sociodemographic and clinical characteristics (eg, age and sex, date of diagnosis, cancer type, and primary treatment). Treating hemato-oncologists were asked to verify the eligibility of the patients. As we aimed to keep a population-based approach, we only defined a few exclusion criteria, such as presence of severe psychopathology or dementia, being in transition to terminal care, and not being able to complete a Dutch questionnaire. Patients were informed that completion of the web-based questionnaire resulted in RCT enrollment with automatic randomization to 1 of the 3 study arms. Paper respondents were not eligible for the RCT (as the return of PRO results and *Living with lymphoma* were web-based) and were observationally followed within the PROFILES lymphoma registry. A reminder mail was sent after 3 weeks. Respondents received follow-up questionnaires at 4 months (T1), 12 months (T2), and 24 months (T3) after the baseline questionnaire. More details about the study design, enrollment, and sample size calculation have previously been published in the research protocol [24]. The RCT was centrally and locally approved by a medical research ethics committee [24].

Randomization

Randomization was performed using block randomization to ensure a balance in sample size across arms over time [27]. Participants were equally randomized to (1) care as usual (CAU), (2) CAU plus return of PRO results, or (3) CAU plus return of PRO results and Living with lymphoma.

Interventions Versus CAU

Arm 1: CAU

In arm 1, patients received CAU from their hemato-oncologists and oncology nurses. In general, they provided verbal information to their patients and provided leaflets regarding the diagnosis and treatment they received.

Arm 2: Return of PRO Results

In arms 2 and 3, in addition to CAU, individual PRO research results were disclosed to patients. This feedback was automatically generated after the completion of the questionnaire. On the basis of respect for autonomy, patients had the choice as to whether they wanted to receive the results [14] and could click on the *feedback* button for return of results.

Detailed information about the return of PRO results (also known as PRO feedback) has been described elsewhere [15,24]. In short, individual research results on general health-related quality of life (HRQoL), physical, emotional, cognitive, and social functioning, fatigue, neuropathy, anxiety, and depressive symptoms were returned to patients [24]. Individual scores were integrated into graphical displays with colored bar charts [28,29]. Patients had the opportunity to compare their scores to mean scores of other patients with lymphoma and an age- and sex-matched normative population without cancer [30] to determine whether their scores were average or not. The colors of the bar charts were related to clinically relevantly mean differences of the evidence-based guidelines of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Organization [31] and considered *average* (amber); *above average* (green); or *below average* (red). Patients with above-average symptom scores were advised to contact their general practitioner.

Arm 3: Return of PRO Results + Living With Lymphoma

In addition to the return of PRO results, patients in arm 3 had access to a web-based self-management intervention—Living with lymphoma. A detailed description of Living with lymphoma has been described elsewhere [24]. Living with lymphoma, an adaptation of the evidence-based BREATH (Breast Cancer eHealth) intervention for breast cancer [32,33], was based on psychoeducation and cognitive behavioral therapy techniques to enhance patients' knowledge and skills. The intervention also included a library with background and additional information on various subjects (eg, work, sexuality, lifestyle) and reference to additional health care services (eg, psychologists, physiotherapists). It was left to the discretion of the patients how and to what extent they used the intervention. The intervention was fully automated, nonguided, and delivered without professional therapist support.

Measures

Sociodemographic and Clinical Measures

Sociodemographic characteristics (age and sex) and detailed clinical information (date of diagnosis, cancer type, and primary treatment) were obtained from the NCR. NCR data were available for both RCT participants and nonparticipants. Information on educational level and marital status was assessed in the questionnaire (data only available for RCT participants).

Comorbidities at the time of questionnaire completion were assessed using an adapted version of the Self-Administered Comorbidity Questionnaire [34]. Patients were asked to identify comorbidities present within the past 12 months—heart disease, hypertension, arthritis, stroke, lung disease, diabetes, stomach disease, kidney disease, liver disease, anemia, thyroid disease, and rheumatoid arthritis. Positive responses were summed to a total score ranging from 0 to 12 (data only available for RCT participants).

Psychological Measures

Personality traits were assessed using the Big Five Inventory [35], a 44-item inventory for measuring the Big Five personality traits—neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Items are scored on a 5-point scale. Scale scores were obtained by averaging all items for each trait and range from 0 to 5 [36].

The 40-item Mental Adjustment to Cancer scale was used to assess adjustment to cancer in terms of coping strategies [37,38]. Items were grouped on five categories: *Helplessness/Hopelessness*, *Anxious Preoccupation*, *Fighting*

Spirit, *Fatalism*, and *Avoidance*. Each item is rated on a 4-point scale. Scale scores were obtained by averaging all items of each strategy and range from 1 to 4. Higher scores represent higher endorsement of the coping strategy.

Health-related quality of life was assessed using the 30-item Quality of Life Questionnaire from the European Organisation for Research and Treatment of Cancer [39]. This questionnaire includes five functional scales, three symptom scales, a global health and quality of life scale, and several single-item symptom measures. Items are scored on a 4-point Likert scale, except for the global health and quality of life scale that is scored on a 7-point linear analog scale. After linear transformation, all scales and single item measures range in score from 0 to 100. Higher scores on functional and health and quality of life scales indicate better functioning or HRQoL, whereas higher scores on symptom scales indicate more complaints.

Self-management Skills

Self-management skills were assessed using the Health Education Impact Questionnaire, which contains 40 items across 8 scales—positive and active engagement in life, health-directed activities, skill and technique acquisition, constructive attitudes and approaches, self-monitoring and insight, health service navigation, social integration and support, and emotional distress [40]. Each item was scored on a 4-point scale. Scale scores were obtained by averaging all items of each domain and ranged from 1 to 4. Higher scores indicate better status or self-management, except for emotional distress, in which higher scores indicate higher distress [40].

Satisfaction With Information Provision

Satisfaction with overall information provision was assessed using an adapted version of the 9-item Information Satisfaction Questionnaire [41]. Patients were also asked to rate their level of satisfaction with overall information on a scale that ranged from 1 (“very unsatisfied”) to 5 (“very satisfied”). In addition, information on patient information preferences was available. Patients were asked to categorize themselves into those who would like (1) all available information, (2) only positive information, and (3) only limited information about their disease. Finally, patients had to categorize themselves into those who would like (1) to be involved in the decision-making process about their disease, or (2) the physician to make the decisions.

Psychological Distress

Psychological distress was assessed using the 14-item Hospital Anxiety and Depression Scale [42]. Each item was rated on a 4-point scale ranging from 0 to 3. The sum score was obtained by adding all item scores and ranged from 0 to 42. Higher scores indicated higher levels of psychological distress [43]. Patients with a Hospital Anxiety and Depression Scale sum score ≥13 were categorized as *psychologically distressed* [44].

Health Care Use

Two open questions were asked to assess health care use: (1) “How often did you contact a general practitioner in the past 12 months?” and (2) “How often did you visit a medical specialist in the past 12 months?”

Statistical Analyses

Overview

Sociodemographic and clinical characteristics between the three arms were compared using univariable analyses of variance and chi-square tests. If at least half of the items from a subscale were completed, the missing items were replaced by the average of those that were present for the participant.

To model between-group differences in change from baseline (T0) to follow-up (T1-T3), mixed-effects models were used with an unstructured covariance structure and a restricted maximum likelihood solution [45]. A random intercept at the patient level was included to adjust for interdependency between repeated measures. The CAU-arm (arm 1) was assigned as the reference group. The *P* value for overall model effects was set at .05, and for specific contrasts at .01, lowering the risk of type I errors as a result of multiple testing. In the iterative process of variable selection, a priori selected covariates (age, sex, cancer type, and treatment) were removed from the model as they were nonsignificant and had no confounders. However, as those in the CAU arm seemed to be somewhat more often psychologically distressed (17/77, 22%) than patients in the return of results arm (9/74, 12%) and the arm with return of results and access to Living with lymphoma (8/76, 11%; *P*=.10; Table 1), we considered psychological distress as a confounding factor and adjusted for baseline psychological distress in analyses (when psychological distress was not the outcome variable). Group differences in mean change scores from baseline to follow-up were accompanied by Cohen effect size (ES). Cohen ES was calculated by dividing the difference in mean change scores between the control and intervention groups by the pooled baseline SD. An ES of 0.20 was considered small, 0.50 moderate, and 0.80 large [44,46]. All analyses were conducted on an intention-to-treat basis. All statistical analyses were performed using SAS version 9.4.

Table 1. Baseline characteristics of participants according to randomized controlled trial (RCT) arm and of nonparticipants.

	CAU[a] (n=77)	Return of PRO[b] results (n=74)	Return of PRO results + living with lymphoma (n=76)	*P* value[c]	Total RCT participants (n=227)	Nonparticipants (n=666)	*P* value[d]
Sociodemographic characteristics							
Age at time of questionnaire (years), mean (SD)	61.3 (12.9)	60.0 (13.4)	60.8 (14.0)	.83	60.7 (13.4)	65.3 (15.7)	<.001
Sex, n (%)				.73			<.001
Male	53 (68.8)	55 (74.3)	53 (69.7)		161 (70.9)	377 (56.6)	
Female	24 (31.2)	19 (25.7)	23 (30.3)		66 (29.1)	289 (43.4)	
Educational level[e], n (%)				.71			—[g]
Low	1 (1.3)	2 (2.7)	3 (3.9)		6 (2.6)	N/A[f]	
Medium	33 (42.9)	38 (51.4)	35 (46.1)		106 (46.7)	N/A	
High	42 (54.5)	34 (45.9)	38 (50)		114 (50.2)	N/A	
Partner (yes), n (%)	62 (80.5)	65 (87.8)	63 (82.9)	.46	190 (83.7)	N/A	—
Clinical characteristics							
Months since diagnosis: mean (SD)	14.0 (3.1)	14.0 (3.6)	13.9 (3.0)	.94	14.0 (3.2)	14.0 (3.5)	.80
Cancer type, n (%)							
Hodgkin lymphoma	10 (12.9)	8 (10.8)	9 (11.8)	.99	27 (11.9)	75 (11.3)	.95
NHL-HG[h]	41 (53.2)	41 (55.4)	42 (55.3)		125 (55.1)	359 (53.9)	
NHL-LG[i]	18 (23.4)	19 (25.7)	19 (25)		56 (24.7)	169 (25.4)	
CLL[j]	8 (10.4)	6 (8.1)	6 (7.9)		19 (8.4)	63 (9.5)	
Primary treatment, n (%)				.10			.13
Active surveillance	23 (29.9)	18 (24.3)	12 (15.8)		53 (23.3)	199 (28.5)	
Received active treatment	52 (67.5)	56 (75.7)	64 (84.2)		172 (75.8)	458 (68.8)	
Chemotherapy	37 (48.1)	46 (62.2)	54 (71.1)		137 (60.4)	357 (54)	
Radiotherapy	5 (6.5)	4 (5.4)	3 (3.9)		12 (5.3)	60 (8.1)	
Stem cell transplantation	8 (10.4)	3 (4.1)	3 (3.9)		14 (6.2)	4 (0.6)	
Other	2 (2.6)	3 (4.1)	4 (5.3)		9 (3.9)	37 (5.6)	
Ann Arbor stage				.81			.30
Stage I, n (%)	6 (7.8)	8 (10.8)	15 (19.7)		29 (12.8)	112 (16.8)	
Stage II, n (%)	8 (10.4)	13 (17.6)	15 (19.7)		36 (15.8)	95 (14.3)	
Stage III, n (%)	11 (14.3)	9 (12.2)	10 (13.2)		30 (13.2)	97 (14.6)	
Stage IV, n (%)	30 (38.9)	30 (40.5)	26 (34.2)		86 (37.9)	194 (29.1)	
Not determined (CLL) or missing, n (%)	22 (28.6)	14 (18.9)	10 (13.2)		46 (20.3)	168 (25.2)	
Number of self-reported comorbidities: mean (SD)	1.3 (1.2)	1.2 (1.1)	1.0 (1.0)	.22	1.1 (1.1)	N/A	
Psychological characteristics, n (%)							
Psychological distress	17 (22.1)	9 (12.2)	8 (10.5)	.09	34 (14.9)	N/A	—

[a]CAU: care as usual.

[b]PRO: patient-reported outcome.

[c]Reports comparisons between the intervention arms and the care as usual arm.

[d]Reports comparisons between randomized controlled trial participants and nonparticipants.

[e]Education levels were low=none or primary school; medium=lower general secondary education or vocational training; or high=preuniversity education

or high-level vocational training or university.

[f]N/A: not applicable.

[g]Not available.

[h]NHL-HG: high-grade non-Hodgkin lymphoma.

[i]NHL-LG: low-grade non-Hodgkin lymphoma.

[j]CLL: chronic lymphocytic leukemia.

Power Calculation

With more than 74 participants per study arm, the study had 90% power to detect an ES of 0.50 with a two-tailed *P* value set at .05 [47].

Results

Baseline Characteristics of RCT Participants

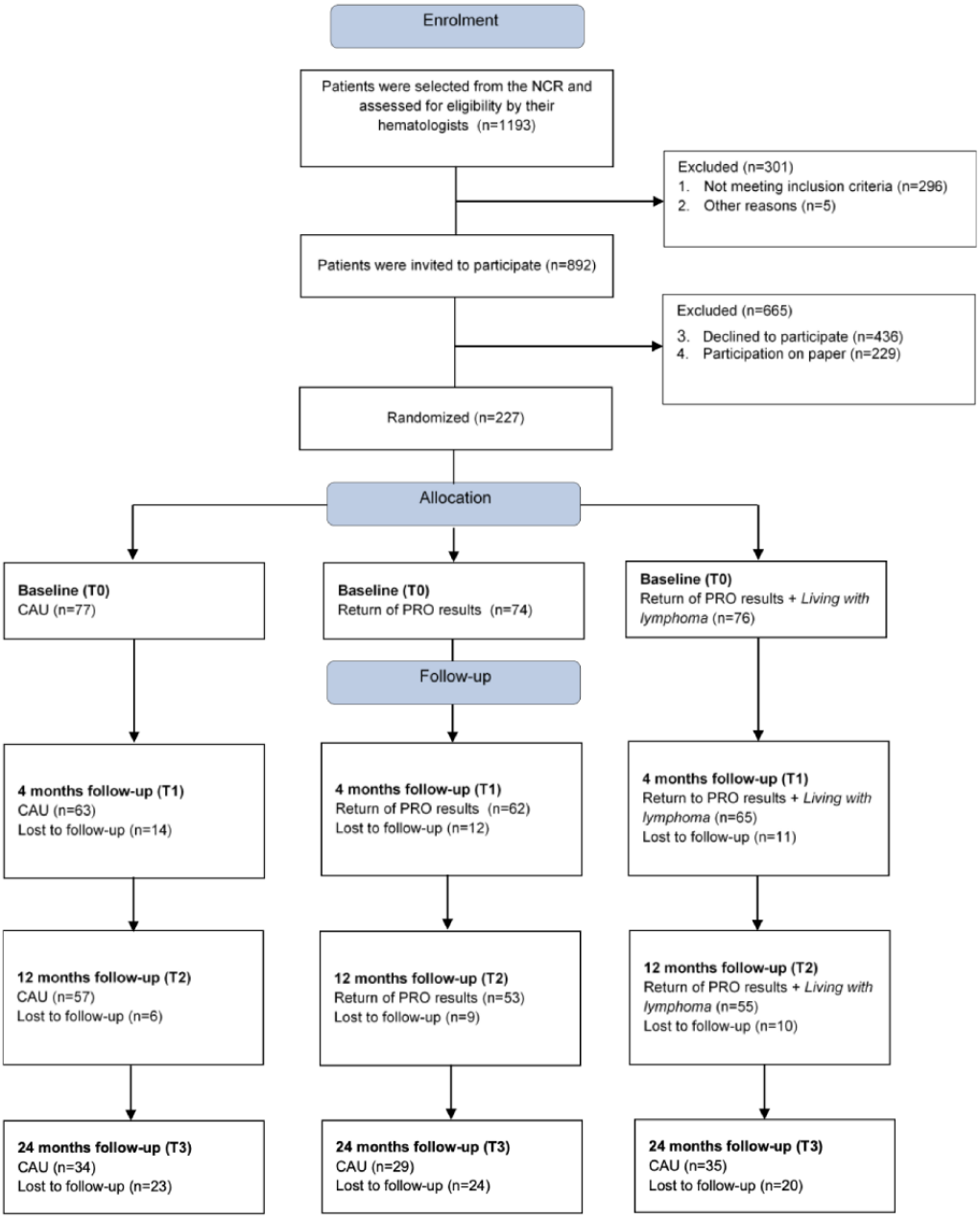
In total, 1193 patients were selected from the NCR and 892 patients were invited to participate. Of these, 51.1% (456/892) participated, of which 25.7% (229/892) were excluded for the RCT as they completed the questionnaire on paper (CONSORT [Consolidated Standards of Reporting Trials] diagram; Figure 1).

Overall, 25.4% (227/892) completed the web-based questionnaire and were included in the RCT. They were randomly assigned to CAU (control group; n=76), return of PRO results (n=74), or return of PRO results and access to Living with lymphoma (n=77). Completion rates of follow-up questionnaires were 84.1% (191/227) on T1, 74.9% (143/191) on T2, and 68.5% (98/143) on T3, and did not differ significantly among groups.

Those who declined participation or completed the questionnaire on paper were analyzed as nonparticipants in this study. All participants provided written informed consent.

RCT participants were younger than nonparticipants (60.7 vs 65.3 years; *P*<.001), and more often men (161/227, 70.9% vs 377/666, 56.6%; *P*<.001). RCT participants were on average 14.0 months after diagnosis (SD 3.2 months). The majority of RCT participants had a partner (190/227, 83.7%). Of the RCT participants, 75.8% (172/227) received active treatment, mostly chemotherapy (137/227, 60.4%), whereas 23.3% (53/227) were on active surveillance. The majority of patients had Ann Arbor stage IV disease at the time of diagnosis. Other baseline sociodemographic and clinical characteristics did not differ across groups (Table 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the progress of the patients with lymphoma through the phases of the Lymphoma Intervention (LIVE) trial. CAU: care as usual (control group); NCR: Netherlands Cancer Registry; PRO: patient-reported outcome.



The Living With Lymphoma Self-management Intervention

The use of self-management intervention was very low (3/76, 4%). Therefore, the effects of *Living with lymphoma* could not be determined within this RCT and were not included in the results. The analyses were performed with the three original RCT arms to maintain a power of 90%.

As we observed that adherence to Living with lymphoma intervention was very low one year after starting patient inclusion, research assistants sent an email for inquiry. A random sample of 5 patients who had access to the intervention and signed up were asked to respond without any obligation and were asked for their reasons for nonadherence. Two patients responded that they felt well and still had regular appointments with their hematologist, and therefore were not in need of an intervention:

> *In the first place, I feel very well, both physically and mentally. Secondly, I have regular appointments with my treating haematologist. Furthermore, I do not always want to be confronted with my disease.* [Male, 79 years]

> *I am not really concerned about the fact that I have had cancer. I am in remission for one year now, and that is how I feel. Every three months, I still have checkup appointments, but other than that I live my life the way I did before I had cancer. I feel fine, I have no limitations, and therefore I do not need information about a disease from the past.* [Male, 54 years]

Return of Individual PRO Results

No statistically significant differences were observed for psychological distress, self-management subscales, and satisfaction with information provision between patients who received their individual PRO results and those who did not (Table 2). In addition, the return of PRO results did not have a significant effect on health care use. As no significant overall group-by-time interaction was found for the outcome variables, we were not allowed to explore specific contrasts.

Table 2. Between-group differences in mean change from baseline to follow-up.

	T0		T1		T2		T3		Between-group difference, T0-T1		Between-group difference, T0-T2		Between-group difference, T0-T3	
	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value
Psychological distress (HADS[a] total; *P*=.94[b])														
CAU[c]	77	7.03 (7.10)	62	6.98 (7.48)	54	6.06 (6.09)	34	6.94 (6.70)	N/A[d]	N/A	N/A	N/A	N/A	N/A
Return of PRO[e] results	74	6.59 (5.26)	60	6.43 (6.21)	53	6.18 (5.61)	28	7.11 (8.26)	0.04 (0.68)	.96	0.20 (0.71)	.77	–0.71 (0.88)	.42
Return of PRO results + *Living with lymphoma*	76	5.75 (5.04)	64	5.78 (4.99)	59	5.68 (5.62)	35	5.34 (5.78)	0.02 (0.67)	.98	0.12 (0.69)	.86	–0.87 (0.83)	.30
Self-management skills (HeiQ positive and active engagement in life[f]; *P*=.52[b])														
CAU	76	3.16 (0.52)	62	3.15 (0.52)	53	3.25 (0.49)	34	3.16 (0.48)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.22 (0.46)	62	2.23 (0.50)	53	3.23 (0.51)	29	3.17 (0.65)	0.02 (0.07)	.81	–0.11 (0.07)	.12	–0.04 (0.09)	.66
Return of PRO results + *Living with lymphoma*	76	3.22 (0.44)	64	3.26 (0.52)	59	3.22 (0.48)	35	3.24 (0.41)	0.07 (0.07)	.34	–0.07 (0.07)	.33	0.02 (0.09)	.82
HeiQ health-directed behavior (*P*=.80[b])														
CAU	77	3.34 (0.62)	62	3.28 (0.56)	55	3.35 (0.64)	34	3.16 (0.57)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.27 (0.53)	62	3.29 (0.59)	53	3.28 (0.49)	29	3.19 (0.47)	0.08 (0.09)	.33	0.00 (0.09)	.69	0.15 (0.11)	.19
Return of PRO results + *Living with lymphoma*	76	3.22 (0.58)	64	3.23 (0.65)	59	3.26 (0.63)	35	3.21 (0.66)	0.09 (0.09)	.27	0.04 (0.09)	.97	0.12 (0.11)	.28
HeiQ skill and technique acquisition (*P*=.90[b])														
CAU	76	2.94 (0.53)	62	2.97 (0.46)	53	3.02 (0.49)	34	3.05 (0.45)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	2.94 (0.44)	61	2.98 (0.43)	53	3.05 (0.49)	29	3.04 (0.55)	0.01 (0.08)	.88	–0.00 (0.08)	.97	0.00 (0.10)	.96
Return of PRO results + *Living with lymphoma*	76	2.97 (0.53)	64	2.95 (0.50)	58	2.98 (0.47)	35	2.94 (0.40)	-0.05 (0.07)	.53	–0.05 (0.08)	.50	–0.11 (0.09)	.22
HeiQ constructive attitudes and approaches (*P*=.36[b])														
CAU	76	3.26 (0.52)	62	3.16 (0.51)	55	3.31 (0.50)	34	3.29 (0.47)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.27 (0.51)	61	3.30 (0.49)	53	3.35 (0.44)	29	3.26 (0.60)	0.13 (0.07)	.06	–0.01 (0.07)	.93	0.03 (0.09)	.73
Return of PRO results + *Living with lymphoma*	76	3.33 (0.43)	64	3.30 (0.49)	58	3.31 (0.45)	35	3.34 (0.44)	0.07 (0.07)	.28	–0.06 (0.07)	.41	–0.06 (0.09)	.52
HeiQ self-monitoring and insight (*P*=.82[b])														
CAU	77	3.08 (0.45)	62	3.17 (0.32)	55	3.15 (0.40)	34	3.04 (0.53)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.05 (0.39)	62	3.11 (0.40)	53	3.13 (0.42)	29	3.14 (0.35)	0.06 (0.07)	.36	–0.02 (0.07)	.78	0.09 (0.09)	.31

	T0		T1		T2		T3		Between-group difference, T0-T1		Between-group difference, T0-T2		Between-group difference, T0-T3	
	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value
Return of PRO results + *Living with lymphoma*	76	3.00 (0.42)	64	3.05 (0.38)	59	3.11 (0.39)	35	3.05 (0.40)	0.07 (0.07)	.30	0.02 (0.07)	.75	0.08 (0.08)	.33
HeiQ health services navigation (*P*=.26[b])														
CAU	76	3.30 (0.45)	62	3.22 (0.44)	55	3.30 (0.47)	34	3.29 (0.42)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.28 (0.39)	61	3.26 (0.40)	53	3.38 (0.43)	29	3.33 (0.44)	0.06 (0.06)	.35	0.07 (0.07)	.30	0.04 (0.08)	.61
Return of PRO results + *Living with lymphoma*	76	3.22 (0.49)	64	3.19 (0.50)	58	3.22 (0.41)	35	3.15 (0.36)	0.05 (0.06)	.40	0.02 (0.07)	.75	–0.13 (0.08)	.11
HeiQ social integration and support (*P*=.68[b])														
CAU	76	3.15 (0.57)	62	3.07 (0.54)	55	3.12 (0.51)	34	3.13 (0.50)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.12 (0.46)	61	3.13 (0.50)	53	3.15 (0.47)	29	3.16 (0.41)	0.06 (0.07)	.38	0.02 (0.07)	.82	0.04 (0.090)	.65
Return of PRO results + *Living with lymphoma*	76	3.22 (0.47)	64	3.20 (0.46)	58	3.16 (0.49)	35	3.19 (0.38)	0.06 (0.07)	.38	-0.02 (0.07)	.76	–0.08 (0.09)	.33
HeiQ emotional wellbeing (*P*=.90[b])														
CAU	76	1.82 (0.62)	62	1.88 (0.65)	55	1.72 (0.55)	34	1.80 (0.50)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	1.83 (0.52)	62	1.82 (0.53)	53	1.73 (0.53)	29	1.88 (0.68)	–0.06 (0.07)	.45	0.02 (0.08)	.76	–0.02 (0.10)	.81
Return of PRO results + *Living with lymphoma*	76	1.71 (0.48)	64	1.72 (0.51)	59	1.65 (0.49)	35	1.64 (0.53)	–0.05 (0.07)	.53	0.02 (0.08)	.82	–0.08 (0.09)	.36
Satisfaction with information provision (ISF[g] total information provision; *P*=.66[b])														
CAU	76	3.86 (0.78)	63	3.73 (0.79)	57	3.88 (0.76)	34	3.85 (0.82)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	3.91 (0.72)	62	3.95 (0.71)	53	4.02 (0.57)	28	3.79 (0.79)	0.19 (0.13)	.14	0.10 (0.13)	.44	–0.01 (0.16)	.97
Return of PRO results + *Living with lymphoma*	76	3.86 (0.80)	65	3.68 (0.94)	59	3.80 (0.76)	35	3.80 (0.68)	–0.02 (0.12)	.88	–0.02 (0.13)	.85	0.02 (0.15)	.92
Health care use (contacts with general practitioner in past 12 months; *P*=.17[b])														
CAU	74	4.70 (3.92)	63	3.16 (2.57)	57	3.63 (4.51)	34	3.38 (3.09)	N/A	N/A	N/A	N/A	N/A	N/A
Return of PRO results	74	4.78 (4.17)	62	4.40 (4.97)	53	3.68 (5.27)	29	4.90 (6.21)	1.04 (0.66)	.11	–0.15 (0.69)	.83	0.29 (0.85)	.73
Return of PRO results + *Living with lymphoma*	76	4.89 (4.32)	65	3.34 (3.25)	59	3.17 (3.24)	35	2.57 (3.00)	–0.16 (0.65)	.81	–0.97 (0.68)	.15	–1.45 (0.82)	.08
Contacts with medical specialist in past 12 months (*P*=.80[b])														
CAU	76	10.09 (6.27)	63	7.00 (4.12)	56	6.66 (5.92)	34	5.44 (4.54)	N/A	N/A	N/A	N/A	N/A	N/A

	T0		T1		T2		T3		Between-group difference, T0-T1		Between-group difference, T0-T2		Between-group difference, T0-T3	
	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	n	Value, mean (SD)	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value	Value, mean change (SE)	*P* value
Return of PRO results	74	10.42 (6.72)	61	6.70 (4.73)	53	5.70 (5.54)	29	5.21 (5.84)	–0.33 (1.10)	.77	–1.20 (1.15)	.30	–1.59 (1.41)	.26
Return of PRO results + *Living with lymphoma*	74	10.46 (7.18)	63	7.30 (5.43)	58	5.97 (3.79)	35	5.26 (6.34)	0.18 (1.10)	.87	–1.05 (1.14)	.35	–1.52 (1.36)	.26

[a]HADS: Hospital Anxiety and Depression Scale (psychological distress subscale range, 0-42, with higher scores indicating more psychological distress).

[b]*P* value of the overall time-by-group interaction.

[c]CAU: care as usual (control group; reference category).

[d]N/A: not applicable.

[e]PRO: patient-reported outcome (T0, baseline assessment; T1, short-term follow-up assessment at 4 months postrandomization; T2, follow-up assessment at 12 months postrandomization; T3, follow-up assessment at 24 months postrandomization).

[f]HeiQ: Health Education Impact Questionnaire (self-management ability subscales range, 0-4, with higher scores indicating higher levels of self-management ability).

[g]ISF: Information Satisfaction Questionnaire (information satisfaction subscale range, 0-5, with higher scores indicating more satisfaction with perceived information).

Of the 150 patients who were randomized to arm 2 or 3 and had access to individual PRO results, 115 (76.7%) patients viewed their PRO results. The majority (79/115, 68.7%) viewed their PRO results more than once, and 16% (13/79) viewed it more than 5 times.

Patients with lymphoma who viewed their PRO results were more recently diagnosed (13.6 vs 15.0 months; *P*=.03), had a more conscientious personality (3.8 vs 3.6; *P*=.01), and had a less fatalistic coping style than those who did not (2.1 vs 2.2; *P*=.02; Table 3). In addition, patients who did view their PRO results more often wished to receive all available information about the disease compared with those who did not (68/115, 59.1% vs 10/35, 29%; *P*=.004).

Of those who viewed the PRO results, 91.3% (105/115) wished to compare their individual results to other patients with lymphoma and 80.8% (93/115) to a normative population without cancer. Only 7.8% (9/115) solely wanted to view their individual results.

Table 3. Baseline characteristics of patients who viewed their individual patient-reported outcome (PRO) results and those who did not.

	Those who viewed their PRO results (N=115)	Those who did not view their PRO results (N=35)	*P* value[a]
Sociodemographic characteristics			
Age at time of questionnaire (years), mean (SD)	60.1 (13.6)	61.4 (14.0)	.64
Sex, n (%)			.93
Male	83 (72.2)	25 (71.4)	
Female	32 (27.8)	10 (28.6)	
Educational level[b], n (%)			.16
Low	4 (3.5)	1 (2.9)	
Medium	51 (44.3)	22 (62.9)	
High	60 (52.2)	12 (34.3)	
Partner (yes), n (%)	14 (12.2)	8 (22.9)	.12
Clinical characteristics			
Months since diagnosis, mean (SD)	13.6 (3.3)	15.0 (3.3)	.03
Cancer type, n (%)			.21
Hodgkin lymphoma	10 (8.7)	7 (20)	
NHL-HG[c]	67 (58.3)	16 (45.7)	
NHL-LG[d]	30 (26.1)	8 (22.9)	
CLL[e]	8 (6.9)	4 (11.4)	
Treatment, n (%)			.63
Active treatment received	91 (79.1)	29 (82.9)	
Active surveillance	24 (20.9)	6 (17.1)	
Number of self-reported comorbidities, mean (SD)	1.1 (1.1)	1.1 (1.2)	.94
Psychological characteristics			
Personality, mean (SD)			
Openness	3.5 (0.6)	3.4 (0.6)	.43
Conscientiousness	3.8 (0.4)	3.6 (0.5)	.01
Extraversion	3.6 (0.6)	3.5 (0.5)	.66
Agreeableness	3.8 (0.4)	3.7 (0.4)	.14
Neuroticism	2.4 (0.6)	2.4 (0.6)	.78
Coping strategies, mean (SD)			
Fighting spirit	3.0 (0.4)	3.0 (0.4)	.47
Anxious preoccupation	2.3 (0.4)	2.2 (0.4)	.07
Helplessness or hopelessness	1.6 (0.4)	1.6 (0.5)	.87
Fatalism	2.1 (0.4)	2.2 (0.3)	.02
Avoidance	1.7 (0.7)	1.6 (0.6)	.47
Psychological distress (yes), n (%)	101 (87.8)	32 (91.4)	.56
Self-monitoring and insight, mean (SD)	3.1 (0.4)	2.9 (0.4)	.01
Satisfaction with information provision, mean (SD)	3.9 (0.8)	3.9 (0.7)	.76

[a]*P* reports comparisons according to analysis of variance and chi-square tests.

[b]Educational levels were defined as follows: low=none or primary school; medium=lower general secondary education or vocational training; or high=preuniversity education or high-level vocational training or university.

[c]NHL-HG: high-grade non-Hodgkin lymphoma.

[d]NHL-LG: low-grade non-Hodgkin lymphoma.

[e]CLL: chronic lymphocytic leukemia.

Discussion

Principal Findings

The results of this RCT demonstrated that patients were neither in a better nor in a worse situation when their individual PRO results were disclosed, as no effects of return of PRO results were found on psychological distress, self-management skills, and satisfaction with information provision. In addition, patients who received their PRO results did not report more contact with their general practitioner or medical specialist compared with those receiving CAU.

Return of individual PRO research results seems to meet patients' wishes, as the majority of those with access viewed their individual results, of whom two-thirds viewed it more than once. The possibility of comparing their scores with peers was most often chosen, indicating the importance of including normative data to place outcomes in perspective. Almost a quarter chose not to receive their results. Therefore, patients should have the choice as to whether they would like to receive their outcomes [14].

We observed little statistical differences in characteristics of patients who did view their PRO results and those who did not. Patients who viewed their PRO results more frequently wanted to receive all available information about the disease compared with those who did not view their results. This is in line with the literature that patients with a monitoring (information seeking) coping style tend to benefit from more provided information, whereas patients with a blunting (information avoiding) coping style benefit from less information [48,49].

There is increasing interest in integrating the collection of PROs in routine practice to enhance clinical care [50]. Weekly measurement of symptoms by patients during active treatment has proven to be effective, as 2 landmark studies on advanced solid tumors showed improvement in HRQoL and survival [51-53]. In our RCT, return of PRO results took place after treatment completion, including patients without advanced cancer stage, not specifically focusing on symptoms and was returned to patients only and not to their treating physician, which all may have contributed to differences in effects. With the continuing development of new and often expensive therapies for patients with cancer (eg, immunotherapy), further research and monitoring of the early onset and course of symptoms remains highly needed [54].

We could not compare our results to the evidence-based BREATH intervention for breast cancer [32,33], the original web-based self-management intervention from which Living with lymphoma intervention was derived, as uptake in our RCT was too low to study the effect. With respect to the design of the study, it was our explicit intention to examine the uptake and effects of a self-management intervention (without personal contact with a therapist) in a population-based setting (where all patients were asked, and no screening took place) to evaluate if this would be a possibility to provide low-intensity care to the continuously growing group of cancer survivors. This may have contributed to the much lower uptake compared with that of BREATH. Therapist guidance may improve patient engagement with a self-management intervention [55]. An important aspect of engagement is self-efficacy, which is a central element of several therapies, such as behavior change theory and planned behavior as a health access approach.

With respect to the *Living with lymphoma* intervention, we were not able to determine an effect as uptake was too low (3/76, 4%). Only 2 patients opened several components of the cognitive behavioral therapy parts, and various items from the library were viewed: *Reliable information*, *Fatigue*, *Emotional counselling*, and *Nutrition and cancer*, *Exercise*, *Physical counselling*, *Sexuality*, and *Reintegration*. As we observed that adherence was very low, we asked patients about reasons for nonadherence, and they indicated they felt well and still had regular appointments with their hematologist and therefore were not in need of an intervention. Furthermore, the majority of participants were men, and poor engagement in self-management was more common for men, as men may be more reluctant to seek help [56]. In addition, the timing of the intervention may not have been optimal, as participants were on average 14 months after diagnosis and patients appeared more receptive to interventions offered near diagnosis [57]. There is evidence that unguided self-management interventions could be more effective when targeted to those in the greatest need of an intervention, such as patients with low distress [32,58]. This might suggest that the need for intervention in our sample may be low and we may have not reached the right group, despite the invitation of a population-based sample and limited exclusion criteria.

This is the first RCT to study the effect of the return of individual PRO results to patients with lymphoma or to patients with cancer in general. The strengths of this study include the design and linkage of PROs with clinical data from the NCR. Owing to the design of our RCT, within a population-based observational cohort, we had access to sociodemographic information about nonparticipants. This provides a unique opportunity to make clear statements about the representativeness of the sample. Participants of the RCT were younger, more often male, and more often highly educated [59]. Ideally, the RCT sample will be representative of the entire target population so that generalizations about the population can be made. More research is needed to understand why underrepresented patients were not reached and how they could be reached in the future.

Conclusions

In conclusion, the return of individual PRO results seems to meet patients' wishes, even though it had no beneficial effects on patient outcomes; it did not have negative effects either. Therefore, we decided to include and implement the return of individual PRO results in the PROFILES registry. In addition, at this moment, the return of individual PRO results is extended to other cancer types, such as colorectal cancer, and is used in daily clinical practice. No conclusions could be drawn about the effectiveness of the self-management intervention because the uptake was too limited.

Acknowledgments

This randomized controlled trial was financially supported by the Jonker-Driessen Foundation. The Patient-Reported Outcomes Following Initial Treatment and Long-Term Evaluation of Survivorship registry was funded by an investment grant (480-08-009) of the Netherlands Organization for Scientific Research (The Hague, The Netherlands). Both funding agencies had no further role in study design; in the collection, analysis, and interpretation of data; in the writing of this paper; and in the decision to submit the paper for publication.

Authors' Contributions

SO was responsible for data collection, data analysis, and drafting of the manuscript. LPJA was responsible for patient recruitment, data collection, data analysis, and drafting of the manuscript. JMK was responsible for data analysis and contributed to the writing of this manuscript. JP designed the intervention that was originally developed for breast cancer patients and wrote the intervention content. MH, MVDP, AK, CL, WS, DI, JP, MO, RVDG, MN, LT, and EP were responsible for data collection and contributed to writing this manuscript. LVDPF was the project leader and was a major contributor in drafting this manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 709 KB - jmir_v23i12e27886_app1.pdf]

References

1. Behringer K, Goergen H, Müller H, Thielen I, Brillant C, Kreissl S, et al. Cancer-related fatigue in patients with and survivors of Hodgkin lymphoma: the impact on treatment outcome and social reintegration. J Clin Oncol 2016 Dec 20;34(36):4329-4337. [doi: 10.1200/JCO.2016.67.7450] [Medline: 27998235]
2. Daniëls LA, Oerlemans S, Krol AD, Creutzberg CL, van DP. Chronic fatigue in Hodgkin lymphoma survivors and associations with anxiety, depression and comorbidity. Br J Cancer 2014 Feb 18;110(4):868-874 [FREE Full text] [doi: 10.1038/bjc.2013.779] [Medline: 24434433]
3. Oerlemans S, Mols F, Issa DE, Pruijt JH, Peters WG, Lybeert M, et al. A high level of fatigue among long-term survivors of non-Hodgkin's lymphoma: results from the longitudinal population-based PROFILES registry in the south of the Netherlands. Haematologica 2013 Mar;98(3):479-486 [FREE Full text] [doi: 10.3324/haematol.2012.064907] [Medline: 22929981]
4. Williams AM, Zent CS, Janelsins MC. What is known and unknown about chemotherapy-related cognitive impairment in patients with haematological malignancies and areas of needed research. Br J Haematol 2016 Sep;174(6):835-846. [doi: 10.1111/bjh.14211] [Medline: 27391367]
5. Oerlemans S, Mols F, Nijziel MR, Zijlstra WP, Coebergh JW, van de Poll-Franse LV. The course of anxiety and depression for patients with Hodgkin's lymphoma or diffuse large B cell lymphoma: a longitudinal study of the PROFILES registry. J Cancer Surviv 2014 Dec;8(4):555-564. [doi: 10.1007/s11764-014-0367-1] [Medline: 24820429]
6. Faller H, Strahl A, Richard M, Niehues C, Meng K. The prospective relationship between satisfaction with information and symptoms of depression and anxiety in breast cancer: a structural equation modeling analysis. Psychooncology 2017 Nov;26(11):1741-1748. [doi: 10.1002/pon.4358] [Medline: 28024096]
7. Faller H, Koch U, Brähler E, Härter M, Keller M, Schulz H, et al. Satisfaction with information and unmet information needs in men and women with cancer. J Cancer Surviv 2016 Feb;10(1):62-70. [doi: 10.1007/s11764-015-0451-1] [Medline: 25956402]
8. Arts LP, Oerlemans S, Tick L, Koster A, Roerdink HT, van de Poll-Franse LV. More frequent use of health care services among distressed compared with nondistressed survivors of lymphoma and chronic lymphocytic leukemia: results from the population-based PROFILES registry. Cancer 2018 Jul 15;124(14):3016-3024 [FREE Full text] [doi: 10.1002/cncr.31410] [Medline: 29698556]
9. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Qual Life Res 2009 Feb;18(1):115-123. [doi: 10.1007/s11136-008-9430-6] [Medline: 19105048]
10. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di DD, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]

11. Snyder CF, Jensen R, Courtin SO, Wu AW, Website for Outpatient QOL Assessment Research Network. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res 2009 Sep;18(7):793-800 [FREE Full text] [doi: 10.1007/s11136-009-9497-8] [Medline: 19544089]
12. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res 2008 Mar;17(2):179-193. [doi: 10.1007/s11136-007-9295-0] [Medline: 18175207]
13. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
14. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm. Washington, DC: National Academies Press (US); 2018.
15. Oerlemans S, Arts LP, Horevoorts NJ, van de Poll-Franse LV. "Am I normal?" The wishes of patients with lymphoma to compare their patient-reported outcomes with those of their peers. J Med Internet Res 2017 Aug 15;19(8):e288 [FREE Full text] [doi: 10.2196/jmir.7079] [Medline: 28811271]
16. Shalowitz DI, Miller FG. Disclosing individual results of clinical research: implications of respect for participants. JAMA 2005 Aug 10;294(6):737-740. [doi: 10.1001/jama.294.6.737] [Medline: 16091577]
17. Wong CA, Hernandez AF, Califf RM. Return of research results to study participants: uncharted and untested. JAMA 2018 Aug 07;320(5):435-436. [doi: 10.1001/jama.2018.7898] [Medline: 29931289]
18. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Soc Sci Med 2005 Feb;60(4):833-843. [doi: 10.1016/j.socscimed.2004.06.022] [Medline: 15571900]
19. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [Medline: 12401421]
20. 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]
21. Braamse AM, van Meijel B, Visser OJ, Boenink AD, Cuijpers P, Eeltink CE, et al. A randomized clinical trial on the effectiveness of an intervention to treat psychological distress and improve quality of life after autologous stem cell transplantation. Ann Hematol 2016 Jan;95(1):105-114 [FREE Full text] [doi: 10.1007/s00277-015-2509-6] [Medline: 26420062]
22. Foster C, Grimmett C, May CM, Ewings S, Myall M, Hulme C, et al. A web-based intervention (RESTORE) to support self-management of cancer-related fatigue following primary cancer treatment: a multi-centre proof of concept randomised controlled trial. Support Care Cancer 2016 Jun;24(6):2445-2453 [FREE Full text] [doi: 10.1007/s00520-015-3044-7] [Medline: 26643072]
23. Owen JE, Klapow JC, Roth DL, Shuster JL, Bellis J, Meredith R, et al. Randomized pilot of a self-guided internet coping group for women with early-stage breast cancer. Ann Behav Med 2005 Aug;30(1):54-64. [doi: 10.1207/s15324796abm3001_7] [Medline: 16097906]
24. Arts LP, van de Poll-Franse LV, van den Berg SW, Prins JB, Husson O, Mols F, et al. Lymphoma InterVEntion (LIVE) - patient-reported outcome feedback and a web-based self-management intervention for patients with lymphoma: study protocol for a randomised controlled trial. Trials 2017 Apr 28;18(1):199 [FREE Full text] [doi: 10.1186/s13063-017-1943-2] [Medline: 28454583]
25. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, Profiles Registry Group. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. Eur J Cancer 2011 Sep;47(14):2188-2194 [FREE Full text] [doi: 10.1016/j.ejca.2011.04.034] [Medline: 21621408]
26. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin D, et al. International Classification of Diseases for Oncology. 3rd edition. Geneva: World Health Organisation; 2000.
27. Kang M, Ragan BG, Park J. Issues in outcomes research: an overview of randomization techniques for clinical trials. J Athl Train 2008;43(2):215-221 [FREE Full text] [doi: 10.4085/1062-6050-43.2.215] [Medline: 18345348]
28. Brundage M, Feldman-Stewart D, Leis A, Bezjak A, Degner L, Velji K, et al. Communicating quality of life information to cancer patients: a study of six presentation formats. J Clin Oncol 2005 Oct 1;23(28):6949-6956. [doi: 10.1200/JCO.2005.12.514] [Medline: 16192583]
29. Kuijpers W, Giesinger JM, Zabernigg A, Young T, Friend E, Tomaszewska IM, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar;25(3):595-604 [FREE Full text] [doi: 10.1007/s11136-015-1107-3] [Medline: 26353905]
30. van de Poll-Franse LV, Mols F, Gundy CM, Creutzberg CL, Nout RA, Verdonck-de Leeuw IM, et al. Normative data for the EORTC QLQ-C30 and EORTC-sexuality items in the general Dutch population. Eur J Cancer 2011 Mar;47(5):667-675. [doi: 10.1016/j.ejca.2010.11.004] [Medline: 21134739]

31. Cocks K, King MT, Velikova G, Martyn SM, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. J Clin Oncol 2011 Jan 1;29(1):89-96. [doi: 10.1200/JCO.2010.28.0107] [Medline: 21098316]
32. van den Berg SW, Gielissen MF, Custers JA, van der Graaf WT, Ottevanger PB, Prins JB. BREATH: web-based self-management for psychological adjustment after primary breast cancer--results of a multicenter randomized controlled trial. J Clin Oncol 2015 Sep 1;33(25):2763-2771. [doi: 10.1200/JCO.2013.54.9386] [Medline: 26169621]
33. van den Berg SW, Gielissen MF, Ottevanger PB, Prins JB. Rationale of the BREAst cancer e-healTH [BREATH] multicentre randomised controlled trial: an internet-based self-management intervention to foster adjustment after curative breast cancer by decreasing distress and increasing empowerment. BMC Cancer 2012 Sep 07;12:394 [FREE Full text] [doi: 10.1186/1471-2407-12-394] [Medline: 22958799]
34. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003 Apr 15;49(2):156-163 [FREE Full text] [doi: 10.1002/art.10993] [Medline: 12687505]
35. Big Five Inventory. APA PsycTests. 1991. URL: https://doi.apa.org/doiLanding?doi=10.1037%2Ft07550-000 [accessed 2021-09-14]
36. McCrae R, Costa PT. Introduction to the empirical and theoretical status of the five-factor model of personality traits. In: Personality Disorders and the Five-Factor Model of Personality. Washington: American Psychological Association; 2013:15-27.
37. Watson M, Greer S, Young J, Inayat Q, Burgess C, Robertson B. Development of a questionnaire measure of adjustment to cancer: the MAC scale. Psychol Med 1988 Feb;18(1):203-209. [doi: 10.1017/s0033291700002026] [Medline: 3363039]
38. Braeken AP, Kempen GI, Watson M, Houben RM, Gils FC, Lechner L. Psychometric properties of the Dutch version of the Mental Adjustment to Cancer scale in Dutch cancer patients. Psychooncology 2010 Jul;19(7):742-749. [doi: 10.1002/pon.1628] [Medline: 19824025]
39. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
40. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
41. Thomas R, Kaminski E, Stanton E, Williams M. Measuring information strategies in oncology - developing an information satisfaction questionnaire. Eur J Cancer Care (Engl) 2004 Mar;13(1):65-70. [doi: 10.1111/j.1365-2354.2004.00445.x] [Medline: 14961777]
42. 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]
43. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
44. Singer S, Kuhnt S, Götze H, Hauss J, Hinz A, Liebmann A, et al. Hospital anxiety and depression scale cutoff scores for cancer patients in acute care. Br J Cancer 2009 Mar 24;100(6):908-912 [FREE Full text] [doi: 10.1038/sj.bjc.6604952] [Medline: 19240713]
45. Littell RC, Pendergast J, Natarajan R. Modelling covariance structure in the analysis of repeated measures data. Stat Med 2000 Jul 15;19(13):1793-1819. [doi: 10.1002/1097-0258(20000715)19:13<1793::aid-sim482>3.0.co;2-q] [Medline: 10861779]
46. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003 May;41(5):582-592. [doi: 10.1097/01.MLR.0000062554.74615.4C] [Medline: 12719681]
47. Cohen J. Statistical Power Analysis for the Behavioral Sciences Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
48. Miller SM. Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. Cancer 1995 Jul 15;76(2):167-177. [doi: 10.1002/1097-0142(19950715)76:2<167::aid-cncr2820760203>3.0.co;2-k] [Medline: 8625088]
49. de Rooij BH, Ezendam NP, Vos MC, Pijnenborg JM, Boll D, Kruitwagen RF, et al. Patients' information coping styles influence the benefit of a survivorship care plan in the ROGY care trial: new insights for tailored delivery. Cancer 2019 Mar 01;125(5):788-797 [FREE Full text] [doi: 10.1002/cncr.31844] [Medline: 30500067]
50. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
51. Basch E. Patient-reported outcomes: an essential component of oncology drug development and regulatory review. Lancet Oncol 2018 May;19(5):595-597. [doi: 10.1016/S1470-2045(18)30141-4] [Medline: 29726373]

52. 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 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
53. Denis F, Basch E, Septans AL, Bennouna JM, 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]
54. Thanarajasingam G, Minasian LM, Baron F, Cavalli F, De Claro RA, Dueck AC, et al. Beyond maximum grade: modernising the assessment and reporting of adverse events in haematological malignancies. Lancet Haematol 2018 Nov;5(11):e563-e598 [FREE Full text] [doi: 10.1016/S2352-3026(18)30051-6] [Medline: 29907552]
55. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
56. Galdas P, Fell J, Bower P, Kidd L, Blickem C, McPherson K, et al. The effectiveness of self-management support interventions for men with long-term conditions: a systematic review and meta-analysis. BMJ Open 2015 Mar 20;5(3):e006620 [FREE Full text] [doi: 10.1136/bmjopen-2014-006620] [Medline: 25795688]
57. Brebach R, Sharpe L, Costa DS, Rhodes P, Butow P. Psychological intervention targeting distress for cancer patients: a meta-analytic study investigating uptake and adherence. Psychooncology 2016 Aug;25(8):882-890. [doi: 10.1002/pon.4099] [Medline: 26893285]
58. Ugalde A, Haynes K, Boltong A, White V, Krishnasamy M, Schofield P, et al. Self-guided interventions for managing psychological distress in people with cancer - a systematic review. Patient Educ Couns 2017 May;100(5):846-857 [FREE Full text] [doi: 10.1016/j.pec.2016.12.009] [Medline: 28081937]
59. Arts LP, Oerlemans S, Posthuma EF, Issa DE, Oosterveld M, van der Griend R, et al. Web-based self-management for patients with lymphoma: assessment of the reach of intervention of a randomized controlled trial. J Med Internet Res 2020 May 14;22(5):e17018 [FREE Full text] [doi: 10.2196/17018] [Medline: 32406858]

Abbreviations

BREATH: Breast Cancer eHealth
CAU: care as usual
CONSORT: Consolidated Standards of Reporting Trials
ES: effect size
HRQoL: health-related quality of life
LIVE: Lymphoma Intervention
NCR: Netherlands Cancer Registry
PRO: patient-reported outcome
PROFILES: Patient-Reported Outcomes Following Initial Treatment and Long-Term Evaluation of Survivorship
RCT: randomized controlled trial

Edited by R Kukafka; submitted 11.02.21; peer-reviewed by J Lehmann, M Ferrer, L Jibb; comments to author 08.05.21; revised version received 02.07.21; accepted 27.07.21; published 14.12.21.

Please cite as:
Oerlemans S, Arts LPJ, Kieffer JM, Prins J, Hoogendoorn M, van der Poel M, Koster A, Lensen C, Stevens WBC, Issa D, Pruijt JFM, Oosterveld M, van der Griend R, Nijziel M, Tick L, Posthuma EFM, van de Poll-Franse LV
Web-Based Return of Individual Patient-Reported Outcome Results Among Patients With Lymphoma: Randomized Controlled Trial
J Med Internet Res 2021;23(12):e27886
URL: https://www.jmir.org/2021/12/e27886
doi:10.2196/27886
PMID:34904948

Original Paper

Feasibility of a Web-Based Implementation Intervention to Improve Child Dietary Intake in Early Childhood Education and Care: Pilot Randomized Controlled Trial

Courtney Barnes[1,2,3,4], PhD; Sze Lin Yoong[1,2,3,4,5], PhD; Nicole Nathan[1,2,3,4], PhD; Luke Wolfenden[1,2,3,4], PhD; Taya Wedesweiler[1], BHlth PhysEd (Hons); Jayde Kerr[1], BNutrDiet; Dianne S Ward[6,7], EdD; Alice Grady[1,2,3,4], PhD

[1]Hunter New England Population Health, Newcastle, Australia

[2]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[3]Hunter Medical Research Institute, New Lambton, Australia

[4]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

[5]School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

[6]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States

[7]Center for Health Promotion and Disease Prevention, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Courtney Barnes, PhD
Hunter New England Population Health
Locked Bag 10, Wallsend
Newcastle, 2287
Australia
Phone: 61 0249246678
Email: courtney.barnes@health.nsw.gov.au

Abstract

Background: Internationally, the implementation of evidence-based healthy eating policies and practices within early childhood education and care (ECEC) settings that encourage children's healthy diet is recommended. Despite the existence of evidence-based healthy eating practices, research indicates that current implementation rates are inadequate. Web-based approaches provide a potentially effective and less costly approach to support ECEC staff with implementing nutrition policies and practices.

Objective: The broad aim of this pilot randomized controlled trial is to assess the feasibility of assessing the impact of a web-based program together with health promotion officer (HPO) support on ECEC center implementation of healthy eating policies and practices. Specifically, we seek to describe the completion rate of study evaluation processes (participant consent and data collection rates); examine ECEC center uptake, acceptability, and appropriateness of the intervention and implementation strategies; understand the potential cost of delivering and receiving implementation support strategies; and describe the potential impact of the web-based intervention on the implementation of targeted healthy eating practices among centers in the intervention group.

Methods: A 6-month pilot implementation trial using a cluster-randomized controlled trial design was conducted in 22 ECEC centers within the Hunter New England region of New South Wales, Australia. Potentially eligible centers were distributed a recruitment package and telephoned by the research team to assess eligibility and obtain consent. Centers randomly allocated to the intervention group received access to a web-based program, together with HPO support (eg, educational outreach visit and local technical assistance) to implement 5 healthy eating practices. The web-based program incorporated audit with feedback, development of formal implementation blueprints, and educational materials to facilitate improvement in implementation. The centers allocated to the control group received the usual care.

Results: Of the 57 centers approached for the study, 22 (47%) provided consent to participate. Data collection components were completed by 100% (22/22) of the centers. High uptake for implementation strategies provided by HPOs (10/11, 91% to 11/11, 100%) and the web-based program (11/11, 100%) was observed. At follow-up, intervention centers had logged on to the program at an average of 5.18 (SD 2.52) times. The web-based program and implementation support strategies were highly acceptable (10/11, 91% to 11/11, 100%). Implementation of 4 healthy eating practices improved in the intervention group, ranging from 19% (2/11) to 64% (7/11).

Conclusions: This study provides promising pilot data to warrant the conduct of a fully powered implementation trial to assess the impact of the program on ECEC healthy eating practice implementation.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12619001158156; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=378099

International Registered Report Identifier (IRRID): RR2-10.1186/s40814-020-00707-w

(J Med Internet Res 2021;23(12):e25902) doi:10.2196/25902

KEYWORDS

childcare center; web-based; nutrition; healthy eating; randomized controlled trial; intervention; implementation

Introduction

Background

Poor dietary intake in early childhood, including inadequate intake of fruit and vegetables and excessive intake of discretionary foods (high in added sugar, sodium, and saturated fat), is a leading contributor to the development of childhood overweight, obesity, cardiovascular disease, and specific types of cancers [1,2]. Globally, preschool aged children do not meet national dietary recommendations for intake of fruit and vegetable servings, while overconsuming discretionary food items [2-5]. As dietary behaviors developed during childhood are known to track into adulthood [6], population-level interventions (ie, interventions targeting a large proportion of the population) to improve child nutrition are recommended [7,8]. Early childhood education and care (ECEC) is a promising setting for interventions aimed at improving children's nutrition behaviors, as they provide access to a large proportion of children [3,9] for prolonged periods [10] during a crucial period of development [11].

Systematic review evidence has identified numerous ECEC-based interventions effective in improving child nutrition behaviors [12] and center nutrition environments [13], including the implementation of evidence-based ECEC practices associated with improved child dietary intake in care [13,14]. The implementation of such evidence-based practices is recommended within national and international ECEC guidelines and includes the provision of healthy foods, positive educator feeding practices (eg, role modeling healthy food choices), and developing center nutrition policies, which detail center strategies and guidelines to enforce the implementation of healthy eating practices [15-17]. However, despite the existence of such guidelines, numerous studies have indicated that the current implementation of evidence-based healthy eating practices is inadequate [18-21].

A recent Cochrane systematic review identified that multicomponent implementation strategies, including researcher delivered face-to-face nutrition education sessions and ongoing support, can produce small but significant improvements in the implementation of healthy eating practices in ECEC centers [13]. Although potentially effective, there are significant challenges in delivering such interventions at scale (ie, to a large number of ECEC centers), including financial and resource burden on centers and the lack of alignment with center capabilities and infrastructure [13]. Web-based modalities provide a potentially effective and less costly approach to implementing nutrition interventions at scale in this setting. Previous research suggests that the use of such modalities to deliver support to center staff is highly acceptable and fits within the existing center infrastructure (eg, access to computers and internet) [12,22,23]. In addition, these modalities can reach a large proportion of the population [24] and have been associated with improvement in a range of provider behaviors and implementation outcomes in previous research delivered outside the ECEC setting [25,26].

Recent trials examining the impact of web-based interventions on ECEC healthy eating practices have been conducted within menu-based centers (ie, centers that provide food to children). A randomized controlled trial (RCT) conducted in 54 Australian childcare centers evaluated the impact of a web-based menu planning program on center compliance with sector dietary guidelines [27]. Results of the RCT found statistically significant improvements in the servings of core food groups and child diet intake; however, the intervention had nonsignificant improvements in the primary outcome of menu compliance with all food groups. The study reported variable levels of engagement with the web-based program, despite the high uptake of implementation support strategies and high acceptability of the intervention and implementation support provided [27]. In addition, the web-based intervention was deemed a cost-effective alternative to traditional menu planning approaches [23]. Within the United States, a pilot RCT conducted in 31 centers evaluated the impact of the web-based Nutrition and Physical Activity Self-Assessment for Child Care (Go-Nutrition and Physical Activity Self-Assessment for Child Care [Go-NAPSACC]) program on center nutrition environments [28]. Despite improvements in food and beverages provided within intervention centers, no statistically significant differences in center nutrition environments were reported at follow-up [28]. Center engagement with the web-based program was not reported; however, the uptake of the implementation support strategies was high among intervention centers. Findings from the process evaluation indicated that a lack of computer literacy among center staff and the need for additional technical support were barriers to program use [28]. Despite these studies showing promise, no RCTs examining the impact of web-based interventions on ECEC healthy eating practices within lunchbox centers (ie, where parents pack foods for children to consume in care) have been conducted.

Objectives

Given the differences between menu-based and lunchbox centers, there is a need to understand whether such interventions

are feasible in the ECEC setting. Feasibility studies are recommended as they allow researchers to collect data to determine whether an intervention is appropriate for more robust testing and to pilot-test recruitment and data collection methods and tools to inform a larger trial [29]. Thus, the aim of this pilot RCT is to determine the feasibility of conducting a fully powered implementation trial assessing the impact of a web-based program together with health promotion officer (HPO) support, on childcare center implementation of healthy eating policies and practices. Specifically, we seek to (1) describe the completion of study evaluation processes (participant consent and data collection rates); (2) examine ECEC center uptake, acceptability, and appropriateness of the intervention and implementation strategies; (3) understand the potential cost of delivering and receiving the implementation strategies; and (4) describe the potential impact of the web-based intervention on the implementation of healthy eating practices among centers in the intervention group.

Methods

Registration and Ethics Approval

This trial was prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12619001158156) and followed the CONSORT reporting guidelines for pilot and feasibility studies [30]. Ethical approval for the trial was obtained from Hunter New England (HNE; HNE approval 06/07/26/4.04) and the University of Newcastle (approval H-2008-0343) Human Research Ethics Committees.

This trial was originally designed as a cluster RCT using an effectiveness-implementation hybrid type-II design. A hybrid effectiveness-implementation design was used to pilot the potential impact and assess the feasibility of an implementation intervention, while assessing the effectiveness of the intervention in improving child dietary intake in care as described by Curran et al [31]. Owing to COVID-19 precluding center site visits to conduct follow-up data collection, we were unable to undertake child lunchbox and dietary assessments and, as such, have not been reported. Therefore, this paper reports on the pilot implementation outcomes that could still be evaluated at follow-up and were specified in the trial registration and protocol.

Study Design and Setting

A protocol detailing the study design and methodology has been published elsewhere [32]. In brief, a pilot implementation trial using a cluster RCT design was conducted in center-based childcare centers within the HNE region of New South Wales, Australia. The HNE region is socioeconomically and geographically diverse, encompassing metropolitan, regional, and remote communities, with a population of over 920,000 residents [33]. Approximately 422 center-based childcare centers, including preschools and long day care, are located within the HNE region, which typically enroll children aged 0-6 years for an average of 21 hours per week [10,34].

Participant Eligibility and Recruitment

Centers

Centers were eligible to participate in the trial if (1) they enrolled >20 children per day, (2) had internet access, (3) parents provided food for children to consume while attending care (ie, centers did not provide food), (4) they did not participate in any other healthy eating or physical activity intervention, and (5) they were not fully compliant with healthy eating practices (ie, not implementing all 5 practices) specified in the NSW state obesity-prevention program (ie, *Munch & Move*) targeted by the intervention, according to the NSW Ministry of Health data monitoring [35]. Centers were ineligible if they were a mobile preschool or family day care center, did not cater to children aged 2-5 years, catered exclusively for children requiring specialist care, or were classified as an NSW Department of Education center owing to differing operational characteristics.

A list of potentially eligible centers located within the HNE region was obtained from the NSW Ministry of Health [35]. One member of the research team with experience recruiting centers to health promotion trials led the recruitment process and monitored consent rates. First, centers were progressively distributed a recruitment package consisting of a study information statement and consent form in random order. Second, the research team member leading recruitment telephoned centers to discuss study details, assess eligibility, and request consent for study participation [19,36]. The centers continued to be contacted until the required number (n=22) consented. During the telephone call, the research team member also scheduled a 2-day baseline data collection site visit for consenting centers. Recruitment for the study was conducted between August 2019 and October 2019.

Children

For children to be eligible to participate, they were required to (1) have written consent from a parent or guardian, (2) be between the ages of 2 and 5 years, (3) be enrolled to attend the center on at least one of the scheduled days of data collection, and (4) not have a dietary restriction requiring specialized tailoring of their diet (eg, allergies or intellectual or physical disability).

Approximately 2 weeks before the baseline data collection site visit, centers were asked to distribute consent forms and information statements to parents via usual communication methods, including email, communication apps, and child pigeonholes. Trained research assistants with experience in recruitment and data collection attended the childcare centers approximately one week before the site visit and on the days of the site visits to request written consent from parents for their children to participate in the study.

Randomization and Blinding

Following baseline data collection, centers were randomly allocated to the intervention or control group, stratified by center socioeconomic status (SES). On the basis of center postcodes, the 2016 Socio-Economic Indexes for Areas was used to classify centers as being located in the least disadvantaged (high SES) or most disadvantaged (low SES) areas [37]. Center postcodes

ranked in the top 50% of NSW were classified as least disadvantaged and the lower 50% of postcodes as the most disadvantaged. The centers were also stratified by those with a high number of Aboriginal child enrollments (defined as those with >10% Aboriginal child enrollments), in a 1:1 ratio through a block randomization procedure (block sizes 2 or 4) conducted by an independent blinded statistician. Given the nature of the intervention (ie, intervention centers were provided access to a web-based program), the centers were not blinded to group allocation. Data collectors were not blinded to group allocation at follow-up.

Intervention

The intervention aimed to improve the implementation of childcare center–level healthy eating practices. The practices targeted within the intervention are recommended by the NSW state obesity-prevention program (ie, *Munch & Move*) [17] as well as national and international guidelines, acknowledging the association between such practices and improved child dietary intake in care [15,16]. Specifically, the practices included the following:

1. Supporting families to provide healthier foods consistent with dietary guidelines: center staff within the intervention group were provided with healthy eating information and resources via the web-based program and were asked to disseminate these to families via usual center communication methods, such as mobile apps, email, and written information, twice during the intervention period. Center staff were also asked to monitor children's lunchboxes daily for consistency with sector-specific dietary guidelines and provide feedback to parents.
2. Provision of intentional healthy eating learning experiences (eg, gardening and cooking lessons): center staff were asked to provide children with intentional healthy eating learning experiences at least twice per week.
3. Using feeding practices that support children's healthy eating (eg, educator role modeling healthy food choices): center staff were asked to provide encouragement to children to promote healthy eating and trying new foods at every meal and snack occasion. Center staff were also asked to role model consuming healthy food choices and avoid the use of foods to encourage desired behavior.
4. Staff participating in professional development targeting healthy eating: center staff were asked to have at least 50% of the staff to participate in web-based training opportunities specific to staff healthy eating behaviors and center practices.
5. Having a comprehensive written nutrition policy that outlines key healthy eating practices: centers were asked to develop or modify existing nutrition policies to document procedures and strategies to facilitate the implementation of healthy eating practices to improve child diet.

A detailed description of these practices is provided in the study protocol [32].

A web-based program, known as Childcare Electronic Assessment Tool and Support (EATS), was developed by the research team to support center implementation of the 5 targeted healthy eating practices. The centers allocated to the intervention group were provided with free access to the web-based program. The intervention was developed by behavioral science researchers, HPOs, state government representatives, and end users from the ECEC setting, including nominated supervisors and educators.

The Behavior Change Wheel (BCW) [38] was used to guide the development and selection of implementation strategies to support center staff in achieving behavior change. During this process, barriers and enablers to center behavior change identified through a literature review and engagement with ECEC staff and stakeholders were mapped to specific behavior change techniques (BCTs) within the BCW [38]. A suite of implementation strategies, defined according to the expert recommendations for implementing change taxonomy, were then selected to action the BCTs within the intervention [39]. The content and implementation strategies within Childcare EATS were selected to ensure user (ie, center staff) engagement, including self-assessment and action planning components to allow center nominated supervisors to reflect on current practice and housed educational resources to facilitate improvements in staff behavior and center processes. The features of the program were developed to integrate within existing center procedures, (eg, the ability to download feedback from the self-assessment quiz) and national assessment and rating standards (eg, the development of action plans as evidence within quality improvement plans). Extensive pilot testing was undertaken with ECEC staff through face-to-face meetings with HPOs to ensure that the functionality and content of Childcare EATS was appropriate and that any potential barriers to program use were addressed. Limitations from previous web-based interventions conducted within the ECEC setting, including low staff computer literacy, need for ongoing technical support, and competing priorities of ECEC staff were also considered during the development of the program [28,40].

Implementation strategies additional to those embedded within the web-based program identified via the BCW process above were used by HPOs who work within the state local health districts to deliver health promotion initiatives within community-based settings such as childcare centers. The HPOs received a training session and implementation manual before delivering the intervention. In addition, HPOs conducted 2 pilot training sessions, with both internal (health service staff with extensive experience supporting ECEC centers to implement obesity prevention initiatives) and external (ECEC center staff) stakeholders. The application of these implementation strategies within the intervention is summarized in Table 1 using the Proctor framework [41] to enable replication.

Table 1. Implementation strategies and behavior change techniques used within the web-based intervention.

Mode of delivery and implementation strategy according to ERIC[a] [39]	Application of the implementation strategy according to Proctor [41]	Behavior change technique actioned via the implementation strategy
Web-based program		
Audit with feedback	• Actor: web-based program • Action: the Childcare EATS[b] program contained a self-assessment feature for centers to assess implementation of targeted healthy eating practices. Centers were automatically provided with tailored feedback on practice performance. • Target: nominated supervisors and center champion knowledge, behavior and abilities, perceived capabilities, and confidence to implement change • Temporality: commencement of the intervention. Centers were encouraged to complete the self-assessment at least twice during the intervention period to monitor change in practice, following the educational outreach visit. • Dose: twice during the intervention period • Implementation outcome: implementation of healthy eating practices • Justification: provision of feedback on center behavior has been used within previous interventions to facilitate improvement in practice within ECEC[c] centers [28,42]	• Feedback on behavior • Feedback on outcome of behavior • Self-monitoring of behavior
Develop a formal implementation blueprint	• Actor: web-based program • Action: following the completion of self-assessment, centers were encouraged to select goals and develop an action plan within the Childcare EATS program. • Target: nominated supervisors and center champion prioritization and investment and perceived capabilities to implement change; formalized guidance and demonstrated support to implement change • Temporality: commencement of the intervention. Centers were encouraged to develop an action plan at least twice within the intervention period, immediately following the self-assessment (audit with feedback). • Dose: twice during the intervention period • Implementation outcome: implementation of healthy eating practices • Justification: developing a formal implementation blueprint has been used within previous interventions to facilitate improvement in practice within ECEC centers [28].	• Goal-setting (outcome and behavior) • Action planning • Problem solving • Review goals (outcome and behavior)
Distribute educational materials	• Actor: web-based program • Action: the Childcare EATS program housed a suite of materials to assist center implementation of the targeted practices, including fact-sheets and resources to facilitate communication with parents; educational materials to improve staff knowledge; example healthy eating learning experiences; professional development and policy templates. • Target: nominated supervisors and center champions to increase staff member knowledge and abilities to implement practices • Temporality: commencement of the intervention. Centers were encouraged to access resources immediately following action planning (development of a formal implementation blueprint). • Dose: accessed at any time during the intervention period • Implementation outcome: implementation of healthy eating practices • Justification: the provision of support and resources via web-based programs is highly acceptable among ECEC staff and has been used within previous interventions within the ECEC setting [22,27,28].	• Demonstration of behavior • Restructuring the physical environment • Adding objects to the environment • Prompts or cues • Credible source
Health promotion officer		

Mode of delivery and implementation strategy according to ERIC[a] [39]	Application of the implementation strategy according to Proctor [41]	Behavior change technique actioned via the implementation strategy
Educational outreach visit	• Actor: HPO[d] • Action: 1.5-2–hour practical face-to-face training session with an HPO was provided to nominated supervisors and center champions to introduce the web-based program and support implementation of the healthy eating practices. • Target: nominated supervisors and center champion knowledge and ability to implement change • Temporality: one-off face-to-face training session (1.5-2 hours) at the start of the intervention (2-8 weeks after baseline) • Dose: one-off training session • Implementation outcome: adoption of the intervention • Justification: face-to-face training within previous ECEC-based interventions has been highly acceptable and used within previous interventions conducted by the research team [27,42]	• Instruction on how to perform behavior • Demonstration on how to perform behavior
Identify and prepare a center champion	• Actor: center champion • Action: center nominated supervisors were asked to identify and prepare a staff member who could dedicate themselves to endorsing and driving implementation of the intervention within their center and asked to attend the educational outreach visit. • Target: center champions; staff investment and motivation to change, formalized guidance and demonstrated support for staff • Temporality: commencement of the intervention period • Dose: ongoing endorsement and support for use of the web-based program throughout the intervention period • Implementation outcome: adoption of the intervention and implementation of healthy eating practices • Justification: preparing a champion has been identified as an effective strategy to drive implementation and has been used in previous trials by the research team [39,43,44].	• Identification of self as role model • Social support (unspecified)
Mandate change	• Actor: HPO, nominated supervisor, and center champion • Action: an MoU[e] was developed to outline the responsibilities and level of commitment expected from both the center and the HPO in working to implement the targeted healthy eating practices. Center nominated supervisors and champions discussed the MoU with the HPO and tailored the content of the MoU to suit the needs of the center. • Target: nominated supervisors and center champion investment and motivation to change, formalized guidance and demonstrated support for staff • Temporality: MoU drafted during the face-to-face educational outreach visit and finalized and signed by the nominated supervisor, center champion, and HPO 2 weeks following the training • Dose: one-off MoU during the face-to-face educational outreach visit, followed by ongoing advocating and support for use of the web-based program by the nominated supervisor and center champion to center staff during the intervention period • Implementation outcome: adoption of the intervention • Justification: securing executive support from nominated supervisors has been effective in improving implementation of healthy eating practices in previous ECEC-based interventions [19]	• Commitment • Social support (unspecified)
Ongoing consultation and local technical assistance		• Social support (unspecified) • Verbal persuasion about capability

Mode of delivery and implementation strategy according to ERIC[a] [39]	Application of the implementation strategy according to Proctor [41]	Behavior change technique actioned via the implementation strategy
	• Actor: HPO • Action: a telephone call was provided to nominated supervisors and center champions to discuss barriers to center implementation of healthy eating practices and the use of the Childcare EATS program, and to develop strategies to address such barriers. Email and telephone support was provided by HPOs upon center request. • Target: nominated supervisors and center champion prioritization and confidence to implement change, formalized guidance, and support • Temporality: 1 telephone call made to centers approximately 2 months following the face-to-face training session • Dose: once during the intervention period • Implementation outcome: adoption of the intervention and implementation of healthy eating practices • Justification: ongoing consultation has been shown to be effective in improving implementation, staff motivation and problem solving within ECEC-based interventions [45,46].	

[a]ERIC: expert recommendations for implementing change.

[b]EATS: Electronic Assessment Tool and Support.

[c]ECEC: early childhood education and care.

[d]HPO: health promotional officers.

[e]MoU: memorandum of understanding.

Control

Centers allocated to the control group received usual care during the intervention period, including general support from HPOs external to the research team upon request to implement the NSW state obesity-prevention program (ie, *Munch & Move*). The provision of such support was centrally monitored by the research team, with 1 center receiving educational materials to support the implementation of healthy eating and physical activity practices before baseline data collection.

Data Collection and Measures

Baseline data were collected between September 2019 and December 2019, and follow-up data were collected between September 2020 and October 2020. A summary of the study outcomes and time points of measurement is provided in Table 2.

Table 2. Study outcomes and time points of measurement.

Study outcome	Time points of measurement
Center and child demographics	Baseline
Feasibility of the evaluation procedures	
Childcare center and child consent rates	Baseline
Completion of data collection components	Baseline
Uptake, acceptability, and appropriateness of the intervention and implementation strategies	
Delivery of the implementation strategies	6 months
Engagement with the Childcare EATS[a] web-based program	6 months
Acceptability of the implementation strategies	12-month follow-up
Appropriateness of the intervention	12-month follow-up
Cost of implementation strategy delivery	Continuously across study period
Implementation of targeted healthy eating practices within the intervention group	Baseline and 6 months

[a]EATS: Electronic Assessment Tool and Support.

Outcomes: Center and Child Demographics

At baseline, a web-based or telephone interview (depending on center preference) with center nominated supervisors was conducted to collect center demographic information, including the type of center (ie, preschool or long day care), center operating hours, number of Aboriginal or Torres Strait Islander enrollments, and number of children enrolled aged between 2 and 5 years. Center area SES and geographic location were determined using the center postcodes. Nominated supervisor demographic information, including age, was also collected during the baseline interviews. A web-based or telephone

interview (depending on center preference) was conducted with center champions at follow-up to collect demographic information, including age.

Information recorded on parent consent forms was used to examine the child demographics. Parents reported the child's age, sex (as recorded on the child's birth certificate), Aboriginal or Torres Strait Islander background, and usual number of days attending care.

Feasibility of the Evaluation Procedures

The feasibility of the evaluation procedures, defined as the extent to which the research can be effectively carried out within the ECEC setting [47], was assessed via parent and center consent rates and completion of data collection components.

Childcare center and child consent rates were assessed using internal records kept by the research team, center, and child consent forms. Center consent rates were calculated as the number of consenting centers divided by the number of eligible centers that were approached to participate in the study. Reasons for centers declining to participate and ineligibility were recorded by the staff member conducting the recruitment telephone calls. Research assistants present on the days of data collection collated all returned child consent forms, including those from parents who did not provide consent for their child to participate in the study. Class lists specific to the days of data collection were obtained from each participating center to determine the total number of eligible children, with consent rates calculated as the number of consenting children divided by the total number of eligible children.

Completion of data collection components including lunchbox observations and measurements, web-based or telephone interviews with nominated supervisors, and observations of the center nutrition environments, was monitored via internal records kept by the research team. These data collection components were used to evaluate the originally planned trial outcomes related to the center nutrition environment and child dietary intake. Center completion of each individual component of data collection (web-based or telephone interview and assessment of center nutrition environments) was collated and entered into a tracking spreadsheet by a member of the research team. The number of complete child dietary intake data collection forms completed during center site visits was counted and included in the tracking spreadsheet.

Uptake, Acceptability, and Appropriateness of the Intervention and Implementation Strategies

The delivery of the implementation strategies was monitored using internal records maintained by the research team. For each center, the following information was recorded: center receipt of each implementation strategy (ie, number of centers that were offered and accepted or declined each strategy), date, duration, and type (ie, email, telephone, or face-to-face) of each implementation strategy delivered, the role of center staff receiving the implementation strategy (ie, nominated supervisor or center champion), and the delivery of BCTs within each implementation strategy (Table 1).

Engagement with the Childcare EATS web-based program was assessed via Google Analytics [48] embedded within the program. Information collected via the analytics included center completion of self-assessments (ie, audit with feedback), development of action plans (ie, developing a formal implementation blueprint), frequency of centers accessing educational materials, total log-ins to Childcare EATS, and average duration of the log-ins. Such measures have been reported in previous ECEC web-based interventions [27,49].

The acceptability of the implementation strategies, defined as the perception among center staff that the implementation strategies are satisfactory, palatable, or agreeable [47], was assessed through web-based and telephone interviews with nominated supervisors and center champions at follow-up. Interview items were modified from those developed by Weiner et al [50] and those used by the research team in previous ECEC-based studies [27,51]. In total, 10 items captured information on the perceived effectiveness (eg, ease of use and helpful in assessing and improving implementation of practices) of the Childcare EATS web-based program and usefulness of the implementation support strategies [27,47,51]. Nominated supervisors responded to each item on a 5-point Likert scale (1=strongly agree to 5=strongly disagree), with the proportion reporting 2 or lower (agree and strongly agree) for each item calculated.

The appropriateness of the intervention, defined as the perceived fit, relevance, or compatibility of the intervention and for the childcare setting [50], was assessed during the web-based or telephone interview with nominated supervisors at follow-up. In total, 4 items captured information on the perceived fit and suitability of healthy eating practices, using modified items by Weiner et al [50]. Nominated supervisors responded to each item on a 5-point Likert scale (1=strongly agree to 5=strongly disagree), with the proportion reporting 2 or lower (agree and strongly agree) for each item calculated.

Cost to Deliver and Receive Implementation Strategies

The direct cost of each implementation strategy delivered by HPOs, including labor (ie, HPO preparation, administration, and delivery of the strategy) and travel, was calculated. Service delivery costs were recorded by the HPOs delivering the intervention. Costs (in Aus $ and US $, 2019/2020) were calculated by multiplying the time spent (in hours) on each implementation strategy by the hourly wage rate of HPOs delivering the intervention. The cost for nominated supervisors and center champions to receive the implementation strategies delivered by HPOs and embedded within the web-based program was also calculated. Data to calculate center costs were recorded by the HPOs delivering the intervention in addition to the time spent in the web-based program captured by the analytics data. Similar to previous studies examining the cost of receiving interventions within the childcare setting [23], costs were calculated by multiplying the time spent (in hours) receiving each implementation strategy by the estimated hourly wage rate of nominated supervisors and educators [52].

Implementation of Targeted Healthy Eating Practices Within the Intervention Group

Self-reported implementation of the 5 targeted healthy eating practices within the intervention group was assessed via baseline nominated supervisor interview data and self-assessments completed by centers via the web-based program at any time point throughout the intervention. In total, 26 items were based on the validated Environment and Policy Assessment and Observation Self-Report [53] and the tool developed by Dodds et al [54] were used to measure the implementation of the 5 healthy eating practices.

In addition, we also assessed contextual factors influencing the center implementation of healthy eating practices, assessed through web-based and telephone interviews with nominated supervisors at follow-up. A total of 5 interview items were based on constructs within 3 of the 5 domains of the Consolidated Framework for Implementation Research (inner setting: compatibility with center values and direction), innovation characteristics (perceived complexity and cost), and outer setting (external influences such as policies and regulations) to identify factors associated with implementation [55]. Nominated supervisors responded to each item on a 5-point Likert scale (1=strongly agree to 5=strongly disagree), with the proportion reporting 2 or lower (agree and strongly agree) for each item calculated.

Statistical Analysis

All statistical analyses were performed using STATA v14 (StataCorp LLC) [56]. All data were analyzed using descriptive statistics. Chi-square analyses were used to compare characteristics of consenting and nonconsenting centers as well as center and child characteristics between the intervention and control groups at baseline. Center locality was classified as either urban (ie, major cities) or rural (ie, inner regional, outer regional, and remote) according to the Australian Statistical Geography Standard [57]. The 2016 Socio-Economic Indexes for Areas was used to classify centers as being located in the least disadvantaged (high SES) or most disadvantaged (low SES) areas [37]. Center postcodes ranked in the top 50% of NSW were classified as least disadvantaged and the lower 50% of postcodes as the most disadvantaged.

Results

Overview

A total of 22 centers and 448 children participated in the study, with 11 (50%) centers randomized to the intervention group and 11 (50%) to the control group (see Figure 1 for the CONSORT diagram). The demographic characteristics of consenting centers and children are summarized in Table 3. There were no significant differences in center SES or center geographic location between the consenting and nonconsenting centers. In addition, there were no significant differences in the center or child characteristics between the intervention and control groups at baseline.

Figure 1. Study flow diagram.

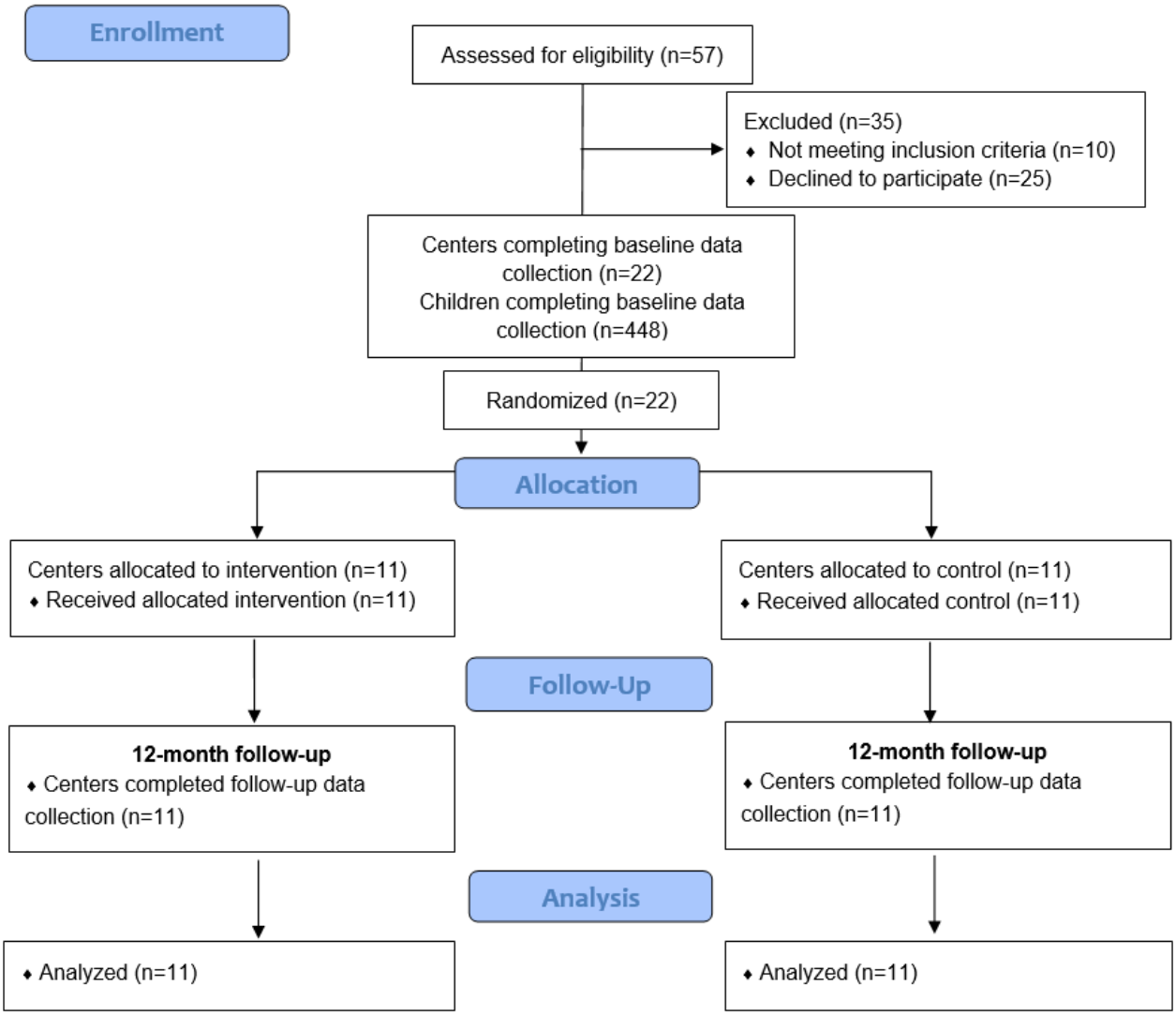


Table 3. Demographic characteristics of participating centers and children (N=11).

Center characteristics	Intervention	Control
Type of center		
Preschool, n (%)	10 (90)	10 (90)
Long day care, n (%)	1 (10)	1 (10)
Child enrollments aged 2-5 years, mean (SD)	30.73 (11.24)	29.0 (8.63)
Aboriginal child enrollments, mean (SD)	5.0 (4.58)	4.64 (3.32)
SEIFA[a,b]		
Most disadvantaged (low SES[c]), n (%)	4 (36)	4 (36)
Least disadvantaged (high SES), n (%)	7 (64)	7 (64)
Geographic location		
Urban (major cities), n (%)	8 (73)	8 (73)
Rural (inner regional, outer regional, and remote), n (%)	3 (27)	3 (27)
Nominated supervisor characteristics		
Age, mean (SD)	37.68 (5.92)	43.91 (10.57)
Center champion characteristics, N	6	—[d]
Age (years), mean (SD)	44.17 (6.40)	
Child characteristics, N	246	202
Age (years), mean (SD)	4.68 (0.66)	4.65 (0.68)
Gender, N	246	202
Female, n (%)	122 (49.5)	88 (43.5)
Male, n (%)	124 (50.4)	114 (56.4)
Children of Aboriginal and Torres Strait Islander background, n (%)	24 (9.7)	20 (9.9)
Days attending care, mean (SD)	2.63 (0.88)	2.57 (0.74)

[a]SEIFA: Socio-Economic Indexes for Areas.

[b]The 2016 Socio-Economic Indexes for Areas was used to classify centers as being located in the least disadvantaged (high socioeconomic status) or most disadvantaged (low socioeconomic status) areas. Center postcodes ranked in the top 50% of New South Wales were classified as least disadvantaged and the lower 50% of postcodes as the most disadvantaged.

[c]SES: socioeconomic status.

[d]Data not available (this item was only applied to nominated supervisors).

Feasibility of the Evaluation Procedures

Childcare Center and Child Consent Rates

Of the 85 potentially eligible centers within the sampling frame, 57 (67%) centers were approached in random order to participate in the study. Of the 57 centers, 10 (18%) centers were ineligible (NSW Department of Education center: 6/10, 60%; involved in another healthy eating or physical activity research trial: 1/10, 10%; and provided food to children: 3/10, 30%) and 25 (44%) centers declined to participate (lack of time: 21/25, 84%; study of lessor importance: 2/25, 8%; and lack of staff capacity: 2/25, 8%). This resulted in an overall study consent rate of 47% (22/57). No centers withdrew from the trial following randomization.

A potential 670 children were eligible to participate in the lunchbox measurements, of whom, 502 (74.9%) provided consent to participate. The consent rate ranged from 53% (16/30) to 96% (24/25) within the participating centers (285/374, 76.2% children within intervention centers and 217/296, 73.3% children within control centers).

Completion of Data Collection Components

Baseline lunchbox observations and measurements conducted to assess the impact of the intervention on child dietary intake were completed for 100% (448/448) of consenting children who were in attendance on data collection days at baseline. The remaining 10.8% (54/502) of the children were absent on the data collection days. Baseline observations of the nutrition environment and web-based or telephone interviews with center nominated supervisors were completed for 100% (22/22) of participating centers.

Uptake, Acceptability, and Appropriateness of the Intervention and Implementation Strategies

Delivery of Implementation Strategies

For implementation strategies delivered by the HPO, 100% of center nominated supervisors or directors were offered and

received the educational outreach visit (ie, face-to-face training session) with the HPO at the commencement of the intervention. The mean duration of the educational outreach visit was 92.73 (SD 21.83) minutes. All centers (n=11) were invited to nominate and prepare a staff member as center champion, with 55% (6/11) of centers nominating a staff member, and 83% (5/6) of these also attending the educational outreach visit. The memorandum of understanding (MoU; ie, mandate change) was drafted with all intervention centers (n=11), with a signed MoU returned by 55% (6/11) of the centers. Ongoing consultation and local technical assistance (ie, follow-up support call provided by the HPO) were offered to 100% (11/11) of the intervention centers, with 91% (10/11) of the centers accepting the call. The mean duration of the follow-up support call was 11.9 (SD 4.70) minutes.

For implementation strategies within the web-based program, overall, 100% (11/11) of centers were provided access to and undertook audit with feedback (ie, self-assessment), developed a formal implementation blueprint (ie, action plan), and accessed the educational materials via the Childcare EATS web-based program.

All intervention centers (n=11) received BCTs as intended in 57% (4/7) of the implementation strategies (Table 4). Additional BCTs (instruction on how to perform the behavior, problem solving, social support [practical], and action planning) were used within the ongoing consultation and local technical assistance strategy in 37% (4/11) of the centers owing to the HPO responding to the needs of the center and tailoring the advice accordingly. Low uptake of the mandate change and identification and preparation of center champion implementation strategies resulted in only 55% (6/11) of the centers receiving the BCTs within these strategies.

Table 4. Behavior change techniques delivered within implementation strategies (N=11).

Mode of delivery, implementation strategy, and behavior change technique	Number of centers
Web-based program	
Audit with feedback	
Feedback on behavior	11 (100)
Feedback on outcome of behavior	11 (100)
Self-monitoring of behavior	11 (100)
Develop a formal implementation blueprint	
Goal-setting (outcome and behavior)	11 (100)
Action planning	11 (100)
Problem solving	11 (100)
Review goals (outcome and behavior)	11 (100)
Distribute educational materials	
Demonstration of behavior	11 (100)
Restructuring the physical environment	11 (100)
Adding objects to the environment	11 (100)
Prompts or cues	11 (100)
Credible source	11 (100)
Health promotion officer	
Educational outreach visit	
Instruction on how to perform behavior	11 (100)
Demonstration on how to perform behavior	11 (100)
Ongoing consultation and local technical assistance	
Social support (unspecified)	10 (91)
Verbal persuasion about capability	10 (91)
Instruction on how to perform behavior[a]	3 (27)
Problem solving[a]	1 (9)
Social support (practical)[a]	1 (9)
Action planning[a]	3 (27)
Mandate change	
Commitment	6 (55)
Social support (unspecified)	6 (55)
Identify and prepare a center champion	
Identification of self as role model	6 (55)
Social support (unspecified)	6 (55)

[a]Additional behavior change techniques used within the ongoing consultation and local technical assistance implementation strategy beyond that specified in the intervention protocol.

Engagement With the Web-Based Program

The intervention center's engagement with the Childcare EATS web-based program is detailed in Table 5. At the 6-month follow-up, intervention centers had logged in to the program on an average of 5.18 (SD 2.52) times, spending an average of 19.90 (SD 11.21) minutes in the program per log-in. Centers completed an average of 2.90 (SD 2.02) self-assessments and developed an average of 2.09 (SD 1.30) action plans. A total of 6 staff members from 4 intervention centers completed web-based professional development accessible via the web-based program or the NSW state obesity-prevention program website (ie, *Munch & Move*) during the intervention period compared with no staff members from control centers.

Table 5. Center engagement with Childcare Electronic Assessment Tool and Support web-based program across 6 months.

Engagement	Value, mean (SD)	Value, median (IQR)
Total log-ins	5.18 (2.52)	4.00 (4.00-5.00)
Average log-in duration (minutes)	19.90 (11.21)	17.44 (10.24-30.03)
Self-assessments completed	2.90 (2.02)	2.00 (1.00-4.00)
Action plans developed	2.09 (1.30)	2.00 (1.00-3.00)
Number of times educational materials were accessed	12.36 (6.71)	10.00 (6.00-18.00)

Acceptability of the Intervention and Implementation Strategies

The web-based program was reported to be an acceptable method for assessing healthy eating practices by most nominated supervisors (10/11, 91%) and center champions (5/6, 83%; Table 6). The implementation strategies provided by HPOs, including the educational outreach visit (ie, face-to-face training) and ongoing support (ie, support call), were considered to be acceptable by nominated supervisors (10/11, 91% to 11/11, 100%). Acceptability of the implementation strategies was lower among center champions (2/6, 33% to 5/6, 83%).

Table 6. Acceptability and appropriateness of the web-based intervention and implementation strategies.

Characteristics	Nominated supervisors (n=11), n (%)	Center champions (n=6), n (%)
Measure (agree or strongly agree)		
Using the web-based program is an acceptable method for assessing if our service is meeting the healthy eating policies and practices.	10 (91)	5 (83)
The web-based program was useful in my service to help meet the healthy eating policies and practices.	11 (100)	5 (83)
Using the web-based program improved my service's performance in meeting the healthy eating policies and practices.	10 (91)	5 (83)
I would recommend the web-based program to other childcare services.	10 (91)	5 (83)
I intend to continue to use the web-based program to help our service meet the healthy eating policies and practices.	10 (91)	5 (83)
I thought the web-based program was easy to use.	10 (91)	—[a]
Measure (useful or very useful)		
I found the face-to-face training session (ie, educational outreach visit) useful.	10 (91)	5 (83)
I found the garnering of managerial support (ie, mandate change) useful.	11 (100)	2 (33)
I found the ongoing telephone support (ie, ongoing consultation and local technical assistance) provided by the health promotion officers useful.	10 (91)	2 (33)
I found nominating a center champion (ie, identify and prepare a center champion) useful.[b]	5 (83)	—
Appropriateness (agree or strongly agree)	11 (100)	—
The healthy eating policies and practices seem fitting.		
The healthy eating policies and practices seems suitable.		
The healthy eating policies and practices seem applicable.		
The healthy eating policies and practices seem like a good match.		
Contextual factors influencing implementation of healthy eating practices (agree or strongly agree)		—
The healthy eating policies and practices are consistent with our center philosophy.	10 (91)	
The healthy eating policies and practices are consistent with the National Quality Framework.	10 (91)	
The healthy eating policies and practices are costly to implement.	0 (0)	
The healthy eating policies and practices are difficult to implement.	4 (36)	
Centers within our region would be supportive of the healthy eating policies and practices.	10 (91)	

[a]Data not available (this item was only applied to nominated supervisors).

[b]This item was only applied to centers that nominated a center champion (n=6).

Appropriateness of the Intervention

In total, 100% (11/11) of nominated supervisors within the intervention group agreed or strongly agreed that healthy eating policies and practices seem fitting, suitable, applicable, and a good match (Table 6).

Cost to Deliver and Receive Implementation Strategies

The total cost to the health service for the HPO to deliver the implementation strategies (ie, educational outreach visit, mandate change, and ongoing consultation) was Aus $ 1351.25 (US $972.64), average per center: Aus $ 122.84 (US $88.42). Overall, the educational outreach visits cost a total of Aus $ 1143.08 (US $822.79), average per center: Aus $ 103.92 (US $74.80), including travel to the center and follow-up correspondence with center staff; mandate change cost a total of Aus $ 43.44 (US $31.27), average per center: Aus $ 3.95 (US $2.84); and ongoing consultation cost a total of Aus $164.73 (US $118.57), average per center: Aus $ 14.98 (US $10.78). The total cost to centers for nominated supervisors and center champions to receive all implementation strategies (ie, those delivered by the HPO and embedded within the web-based program) was Aus $ 1516.40 (US $1091.51), average per center: Aus $ 137.85 (US $99.23). The cost to receive the implementation strategies delivered by the HPO was Aus $ 1052.29 (US $757.44), average per center: Aus $ 95.66 (US $68.86), whereas the cost to receive the implementation strategies embedded within the web-based program was Aus $ 464.11 (US $334.07), average per center: Aus $ 42.19 (US $30.37).

Implementation of Targeted Healthy Eating Practices Within the Intervention Group

The proportion of centers implementing targeted healthy eating practices improved in 4 of the 5 practices from baseline to follow-up (Table 7). The greatest improvement was reported in center educator use of feeding practices that support children's healthy eating, increasing from 18% (2/11) to 82% (9/11). The proportion of centers supporting families to provide healthier foods consistent with dietary guidelines decreased from 82% (9/11) to 55% (6/11). At follow-up, 18% (2/11) of centers were implementing all 5 healthy eating practices, whereas none were at baseline. The mean number of practices implemented per center increased from 3.36 (SD 1.21) at baseline to 4.36 (SD 1.21) at follow-up. When examining the change in practice implementation between the most (low SES) and least (high SES) disadvantaged centers, the number of most disadvantaged centers supporting families to provide healthier foods consistent with dietary guidelines reduced from 100% (4/4) at baseline to 25% (1/4) at follow-up compared with no change in least disadvantaged centers (Table 8).

In total, 91% (10/11) of nominated supervisors reported that healthy eating practices were consistent with the philosophy of their service and consistent with the ECEC settings regulatory standards (ie, the National Quality Framework; Table 6).

Table 7. Intervention group implementation of healthy eating practices (N=11).

Healthy eating practice	Centers implementing at baseline, n (%)	Centers implementing at follow-up, n (%)	Change, n (%)
Provision of intentional healthy eating learning experiences	4 (36)	6 (55)	2 (18)
Comprehensive written nutrition policy that outlines key healthy eating practices	8 (73)	10 (91)	2 (18)
Staff participating in professional development targeting healthy eating	3 (27)	6 (55)	3 (27)
Educator use of feeding practices that support children's healthy eating	2 (18)	9 (82)	7 (64)
Supporting families to provide healthier foods consistent with dietary guidelines	9 (82)	6 (55)	−3 (27)

Table 8. Intervention group implementation of healthy eating practices by Socio-Economic Indexes for Areas classification (N=1)[a].

Healthy eating practice	Low SES[b] (n=4), n (%)			High SES (n=7), n (%)		
	Most disadvantaged centers implementing at baseline	Most disadvantaged centers implementing at follow-up	Change	Least disadvantaged centers implementing at baseline	Least disadvantaged centers implementing at follow-up	Change
Provision of intentional healthy eating learning experiences	2 (50)	2 (50)	0 (0)	2 (29)	4 (57)	2 (29)
Comprehensive written nutrition policy that outlines key healthy eating practices	3 (75)	3 (75)	0 (0)	5 (71)	7 (100)	2 (29)
Staff participating in professional development targeting healthy eating	1 (25)	1 (25)	0 (0)	1 (14)	5 (71)	4 (57)
Educator use of feeding practices that support children's healthy eating	1 (25)	4 (100)	3 (75)	1 (14)	5 (71)	4 (57)
Supporting families to provide healthier foods consistent with dietary guidelines	4 (100)	1 (25)	−3 (75)	5 (71)	5 (71)	0 (0)

[a]The 2016 Socio-Economic Indexes for Areas was used to classify centers as being located in the least disadvantaged (high socioeconomic status) or most disadvantaged (low socioeconomic status) areas. Center postcodes ranked in the top 50% of New South Wales were classified as least disadvantaged and the lower 50% of postcodes as the most disadvantaged.

[b]SES: socioeconomic status.

Discussion

Principal Findings

This study aimed to assess the potential feasibility of a pilot cluster RCT of a web-based healthy eating implementation intervention in ECEC centers to undertake a fully powered implementation trial. The study also examined the uptake, acceptability, appropriateness, and actual cost of delivering the intervention and implementation strategies. Overall, the study findings indicate that the web-based intervention and most implementation strategies are highly feasible, low-cost, and acceptable to childcare center staff and can improve the implementation of healthy eating practices in ECEC centers.

The study obtained a high overall parental consent rate of 74.9% (502/670) for children to participate in lunchbox measurements. However, the variability in parental consent across centers (ranging from 16/30, 53% to 24/25, 96%) is worth noting. This variation may be owing to the differing relationships within centers between staff and parents regarding the contents of children's lunchboxes with previous studies reporting a reluctance from staff to communicate with parents in fear of having difficult conversations [58,59]. As such, some parents may have been reluctant to consent to lunchbox measurements owing to perceived judgment [58,59]. Although not dissimilar to previous web-based studies conducted within the ECEC setting, the overall study consent rate among centers was moderate at 47% [27,60,61]. Similar to previous studies, barriers to center participation reported by staff included a lack of time and competing priorities [62]. As this study attempted to address such barriers by embedding the intervention within usual center processes (ie, aligning with ECEC accreditation standards), further consideration needs to be taken to better promote the intervention by aligning with current center priorities during study recruitment. However, once consented to the trial, the study data collection components were highly feasible, with 100% of participating centers completing child lunchbox measurements, center nutrition environment observations, and interviews with nominated supervisors. This indicates that such methods should be retained for a fully powered implementation trial.

Promising levels of uptake and acceptability of the implementation strategies used in this study were observed. The level of engagement with the web-based program was consistent with recommendations for centers to complete the self-assessment (audit with feedback) and develop action plans (formal implementation blueprint) twice during the intervention period. Such findings suggest that centers are likely to receive the intended dose of the intervention with the current implementation strategies. The promising levels of engagement may be attributed to the web-based program being easy to use as reported by nominated supervisors and aligned with usual center processes [63]. However, large SDs and wide IQRs for the number of log-ins and log-in duration indicate high variability in engagement with the web-based program across centers. Despite such variability being consistent with previous studies within the ECEC setting that used web-based modalities [27], exploration is needed to better understand the reasons behind the relatively lower levels of engagement for some centers.

As the intervention was largely delivered remotely, the overall cost to deliver the implementation strategies was minimal (total of Aus $ 1351.25 [US $ 931.88]; average per center: Aus $ 122.84 [US $ 86]). Therefore, the web-based intervention may be considered a low-cost alternative to support center implementation compared with traditional, highly intensive modalities. However, the study was unable to capture the costs associated with center staff implementing healthy eating practices, including the time spent disseminating information to parents. As such, future studies should consider conducting a cost-effective analysis, while capturing costs associated with center implementation of practices, to enable researchers, practitioners, funding bodies, and centers to determine whether investment in the web-based intervention produced an acceptable return and is a cost-effective approach to support the implementation of healthy eating practices at scale. Consistent with previous studies conducted within the ECEC setting [51,64], high levels of uptake and acceptability were found for most implementation strategies provided by HPOs, particularly the educational outreach visit (11/11, 100%) and local technical assistance (10/11, 91%). Despite previous literature suggesting that implementation strategies such as the MoU and center champions are useful for facilitating the uptake of interventions [19,39,44], the relatively low uptake of these strategies is worth exploring. Although there was high acceptability of the center champion strategy in centers that nominated a champion (9/10, 83%), a potential explanation for the lower uptake of the strategy may be the differing organizational structures within centers. Anecdotally, the uptake of center champions was higher in larger centers with greater staffing numbers and child enrollments, where the nominated supervisor often engages educational leads. The educational lead takes on additional advocacy roles among staff, lending them to the role of the center champion. In smaller centers however, the nominated supervisors often work as the educational lead themselves, acting as the main advocate among center staff. Therefore, the research team should consider alternative strategies, such as a local consensus approach [51] (ie, the entire center), to ensure that the uptake of the intervention remains high in centers where a sole champion is not a feasible strategy.

The improvement in implementation of 4 of the 5 targeted healthy eating practices within the intervention group is promising, with effect sizes ranging from 19% (2/11) to 64% (7/11). Such effect sizes are encouraging when compared with previous studies aimed at improving the implementation of practices within the ECEC setting [13]. A recent Cochrane systematic review, which examined the effectiveness of strategies aimed at improving the implementation of healthy eating and physical activity policies and practices, reported effect sizes as low as 2.5% [13]. Therefore, our findings show great promise for testing in a fully powered implementation trial. However, a decrease in centers supporting families to provide healthier foods consistent with dietary guidelines, particularly in those centers classified as most disadvantaged, is worth noting given this practice had the highest rates of implementation at baseline. A potential explanation for this

reduction may be the competing information relating to COVID-19 distributed to parents during the intervention period (eg, communication regarding center safety protocols and changes to child attendance fees), resulting in support for parents to provide foods consistent with sector dietary guidelines being of lesser priority at this time. Research suggests a lack of skills, knowledge, and confidence in communicating with parents regarding healthy eating [58,59,65] may also negatively impact the implementation of this practice. Using strategies, such as ongoing professional development, coaching, and training, have been suggested in recent studies to address such barriers and support ECEC staff to engage in positive and effective communication with parents [65]. As centers were encouraged to distribute the healthy eating resources to parents via usual communication methods (eg, parent communication apps, email, and written information), further consideration of the most effective method to facilitate staff communication with parents regarding healthy eating and nutrition may be required. Although the Childcare EATS engagement data provided important insights into the center use of the web-based program, the methods used and the reach of the center distribution of healthy eating information and resources (eg, number of parents who received the resources) could not be measured. In addition, we were unable to assess whether parents within the intervention group communicated healthy eating information provided by the center staff to other parents. There was a notable contrast in the implementation of this practice between centers classified as most and least disadvantaged. This contrast may potentially be explained by COVID-19 related impacts on resourcing (eg, staffing, budget, and time) within disadvantaged centers, who may have already been experiencing limited resources before the pandemic. A better understanding of the barriers faced by centers classified as most disadvantaged in communicating with parents should be sought to enable the development of appropriate strategies to support implementation of this practice. However, given the small sample size in this study, this finding is highly exploratory and should be interpreted with caution. In addition, collecting contextual data from parents regarding their preferred method of receiving healthy eating information from centers may also provide guidance on the most effective way to support parents in packing healthy lunchboxes for children to consume in care.

The findings from this study provide compelling data to support the conduct of a fully powered implementation trial. Importantly, despite the relatively low level of support provided to childcare centers to use the program, the level of engagement with the web-based program was relatively high, and large changes in practice implementation were observed. Findings from this study suggest that several improvements could be made to the intervention, including considering the appropriateness of the MoU and center champion and using strategies to support ECEC center staff engagement with parents regarding healthy eating. Finally, the inclusion of a nested evaluation within a future trial to assess the impact of the web-based intervention on individual-level outcomes, including child dietary intake and parent lunchbox packing practices, should be considered to gain greater insight into the effectiveness of the intervention beyond center-level outcomes.

Limitations

Although unavoidable because of restrictions relating to the COVID-19 pandemic, the inability to assess center nutrition environments and conduct child lunchbox assessments via direct observation to assess child-level outcomes as originally intended is a limitation of the study. In addition, although the data regarding the impact of the intervention on center implementation are promising, these data were only able to be collected within intervention centers with no comparison to the control group, and as such, should be interpreted with caution. Finally, as the study was conducted within 1 region of NSW, the generalizability of the findings beyond the region may be limited.

Conclusions

This pilot study provides compelling data to support the conduct of a larger trial assessing the impact of the web-based intervention on ECEC center implementation of healthy eating practices. The findings of this pilot study indicate that the web-based intervention is highly feasible, acceptable, appropriate, and low-cost. As this study is one of few examining the potential impact of a web-based intervention within the ECEC setting, a fully powered implementation trial is warranted to establish the true effects and examine the impact of the intervention at scale.

Acknowledgments

The authors wish to thank the staff at Hunter New England (HNE) Population Health, Population Health Aboriginal Advisory Network Group, and Christophe Lecathelinais for their statistical support. The authors would also like to thank the participating center staff, parents, and children. Finally, the authors would like to thank Kasia Bolsewicz, Melody Ng and Hannah Lamont for their contribution to the planning, implementation, and evaluation of the study. This project was partly funded by the Hunter Children's Research Foundation, Priority Research Center for Health Behavior University of Newcastle, Hunter Cancer Research Alliance, HNE Local Health District Grant, and Hunter Medical Research Institute. The funders played no role in conducting the trial. HNE Population Health and the University of Newcastle provided infrastructure and in-kind funding. CB receives salary support from a cofunded industry scholarship between HNE Population Health and the University of Newcastle. SLY receives salary support from the Australia Research Council Discovery Early Career Research Awards (DE170100382). LW receives salary support from the National Health Medical Research Council Career Development Fellowship (grant APP1128348), Heart Foundation Future Leader Fellowship (grant 101175), and a HNE Clinical Research Fellowship. The contents of this manuscript are the responsibility of the authors and do not reflect the views of Hunter Children's Research Foundation, Hunter Cancer

Research Alliance, Priority Research Center for Health Behavior, Hunter Medical Research Institute Australian Research Council Heart Foundation, HNE Local Health District, or National Health Medical Research Council.

Authors' Contributions

CB drafted the manuscript. SLY, AG, CB, NN, and LW conceived and designed the research and interventions. CB, AG, and JK delivered the intervention. CB, AG, TW, and JK developed the evaluation protocols and led to the acquisition of data. CB conducted the data analyses. All authors contributed to drafting and final approval of the manuscript.

Conflicts of Interest

Authors CB, LW, AG, TW, and JK received salary support from Hunter New England Local Health District, which provided infrastructure and funding for the study.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2954 KB - jmir_v23i12e25902_app1.pdf]

References

1. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
2. Increasing fruit and vegetable consumption to reduce the risk of noncommunicable diseases . World Health Organization. URL: https://www.who.int/elena/titles/fruit_vegetables_ncds/en/ [accessed 2021-11-27]
3. Child care and early childhood education in Australia. Australian Institute of Family Studies. 2015. URL: https://aifs.gov.au/publications/child-care-and-early-childhood-education-australia [accessed 2021-11-27]
4. Fox MK, Gearan E, Cannon J, Briefel R, Deming DM, Eldridge AL, et al. Usual food intakes of 2- and 3-year old U.S. children are not consistent with dietary guidelines. BMC Nutr 2016 Nov 15;2(1):67. [doi: 10.1186/s40795-016-0106-2]
5. NDNS: results from Years 5 and 6 (combined). GOV.UK. URL: https://www.gov.uk/government/statistics/ndns-results-from-years-5-and-6-combined [accessed 2021-11-27]
6. Craigie AM, Lake AA, Kelly SA, Adamson AJ, Mathers JC. Tracking of obesity-related behaviours from childhood to adulthood: a systematic review. Maturitas 2011 Nov;70(3):266-284. [doi: 10.1016/j.maturitas.2011.08.005] [Medline: 21920682]
7. Population-based approaches to childhood obesity prevention. World Health Organization. URL: http://apps.who.int/iris/bitstream/handle/10665/80149/9789241504782_eng.pdf?sequence=1 [accessed 2021-11-27]
8. Recommendations and public health and policy implications. World Cancer Research Fund International. URL: https://www.wcrf.org/wp-content/uploads/2021/01/Recommendations.pdf [accessed 2021-11-27]
9. Hesketh KR, Griffin SJ, van Sluijs EM. UK Preschool-aged children's physical activity levels in childcare and at home: a cross-sectional exploration. Int J Behav Nutr Phys Act 2015 Sep 26;12:123 [FREE Full text] [doi: 10.1186/s12966-015-0286-1] [Medline: 26410252]
10. Childhood education and care, Australia. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/people/education/childhood-education-and-care-australia/latest-release [accessed 2021-11-27]
11. Swinburn B, Malakellis M, Moodie M, Waters E, Gibbs L, Millar L, et al. Large reductions in child overweight and obesity in intervention and comparison communities 3 years after a community project. Pediatr Obes 2014 Dec;9(6):455-462. [doi: 10.1111/j.2047-6310.2013.00201.x] [Medline: 24203373]
12. Matwiejczyk L, Mehta K, Scott J, Tonkin E, Coveney J. Characteristics of effective interventions promoting healthy eating for pre-schoolers in childcare settings: an umbrella review. Nutrients 2018 Mar 01;10(3):293 [FREE Full text] [doi: 10.3390/nu10030293] [Medline: 29494537]
13. Wolfenden L, Barnes C, Jones J, Finch M, Wyse RJ, Kingsland M, et al. Strategies to improve the implementation of healthy eating, physical activity and obesity prevention policies, practices or programmes within childcare services. Cochrane Database Syst Rev 2020 Feb 10;2:CD011779 [FREE Full text] [doi: 10.1002/14651858.CD011779.pub3] [Medline: 32036618]
14. Stacey FG, Finch M, Wolfenden L, Grady A, Jessop K, Wedesweiler T, et al. Evidence of the potential effectiveness of centre-based childcare policies and practices on child diet and physical activity: consolidating evidence from systematic reviews of intervention trials and observational studies. Curr Nutr Rep 2017 Jul 22;6(3):228-246. [doi: 10.1007/s13668-017-0212-z]
15. Benjamin Neelon SE, Briley ME, American Dietetic Association. Position of the American Dietetic Association: benchmarks for nutrition in child care. J Am Diet Assoc 2011 Apr;111(4):607-615. [doi: 10.1016/j.jada.2011.02.016] [Medline: 21443997]

16. Australian Govrenment Department of Health. URL: https://www.health.gov.au/resources/collections/get-up-grow-resource-collection?utm_source=health.gov.au&utm_medium=callout-auto-custom&utm_campaign=digital_transformation [accessed 2021-11-27]
17. Munch & Move. NSW Government. URL: https://healthykids.nsw.gov.au/ [accessed 2021-11-28]
18. Kelly B, Hardy L, Howlett S, King L, Farrell L, Hattersley L. Opening up Australian preschoolers' lunchboxes. Aust N Z J Public Health 2010 Jun;34(3):288-292. [doi: 10.1111/j.1753-6405.2010.00528.x] [Medline: 20618271]
19. Bell AC, Davies L, Finch M, Wolfenden L, Francis JL, Sutherland R, et al. An implementation intervention to encourage healthy eating in centre-based child-care services: impact of the Good for Kids Good for Life programme. Public Health Nutr 2015 Jun;18(9):1610-1619. [doi: 10.1017/S1368980013003364] [Medline: 24477181]
20. Yoong SL, Skelton E, Jones J, Wolfenden L. Do childcare services provide foods in line with the 2013 Australian Dietary guidelines? A cross-sectional study. Aust N Z J Public Health 2014 Dec;38(6):595-596. [doi: 10.1111/1753-6405.12312] [Medline: 25440467]
21. Frampton AM, Sisson SB, Horm D, Campbell JE, Lora K, Ladner JL. What's for lunch? An analysis of lunch menus in 83 urban and rural Oklahoma child-care centers providing all-day care to preschool children. J Acad Nutr Diet 2014 Sep;114(9):1367-1374. [doi: 10.1016/j.jand.2013.09.025] [Medline: 24332085]
22. Wolfenden L, Finch M, Nathan N, Weaver N, Wiggers J, Yoong SL, et al. Factors associated with early childhood education and care service implementation of healthy eating and physical activity policies and practices in Australia: a cross-sectional study. Transl Behav Med 2015 Sep;5(3):327-334 [FREE Full text] [doi: 10.1007/s13142-015-0319-y] [Medline: 26327938]
23. Reeves P, Edmunds K, Szewczyk Z, Grady A, Yoong SL, Wolfenden L, et al. Economic evaluation of a web-based menu planning intervention to improve childcare service adherence with dietary guidelines. Implement Sci 2021 Jan 07;16(1):1 [FREE Full text] [doi: 10.1186/s13012-020-01068-x] [Medline: 33413491]
24. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
25. Schoeppe S, Alley S, Van Lippevelde W, 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]
26. Antwi F, Fazylova N, Garcon M, Lopez L, Rubiano R, Slyer J. The effectiveness of web-based programs on the reduction of childhood obesity in school-aged children: a systematic review. JBI Libr Syst Rev 2012;10(42 Suppl):1-14. [doi: 10.11124/jbisrir-2012-248] [Medline: 27820152]
27. Grady A, Wolfenden L, Wiggers J, Rissel C, Finch M, Flood V, et al. Effectiveness of a web-based menu-planning intervention to improve childcare service compliance with dietary guidelines: randomized controlled trial. J Med Internet Res 2020 Feb 04;22(2):e13401 [FREE Full text] [doi: 10.2196/13401] [Medline: 32014843]
28. Ward DS, Vaughn AE, Mazzucca S, Burney R. Translating a child care based intervention for online delivery: development and randomized pilot study of Go NAPSACC. BMC Public Health 2017 Nov 21;17(1):891 [FREE Full text] [doi: 10.1186/s12889-017-4898-z] [Medline: 29162057]
29. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
30. 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 Oct 21;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
31. Curran G, Bauer M, Mittman B, Pyne J, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
32. Barnes C, Grady A, Nathan N, Wolfenden L, Pond N, McFayden T, et al. A pilot randomised controlled trial of a web-based implementation intervention to increase child intake of fruit and vegetables within childcare centres. Pilot Feasibility Stud 2020 Oct 29;6(1):163 [FREE Full text] [doi: 10.1186/s40814-020-00707-w] [Medline: 33292720]
33. HealthStats NSW. NSW Government. URL: https://www.healthstats.nsw.gov.au/#/home [accessed 2021-11-28]
34. National registers. Australian Children's Education & Care Quality Authority. URL: https://www.acecqa.gov.au/resources/national-registers [accessed 2021-11-28]
35. Population Health Information Management System PHIMS, Healthy Children Initiative. Published 2019. Accessed 18 March 2019.
36. Finch M, Wolfenden L, Morgan PJ, Freund M, Wyse R, Wiggers J. A cluster randomised trial to evaluate a physical activity intervention among 3-5 year old children attending long day care services: study protocol. BMC Public Health 2010 Sep 08;10:534 [FREE Full text] [doi: 10.1186/1471-2458-10-534] [Medline: 20822543]
37. Socio-economic indexes for areas. Australian Bureau of Statistics. URL: https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa [accessed 2021-11-28]
38. 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;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

39. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
40. Wallace R, Costello L, Devine A. Exploring sense of community among early childhood education and care professionals through the Supporting Nutrition for Australian Childcare (SNAC) website. Australasian J Early Childhood 2018 Dec 01;43(4):23-32. [doi: 10.23965/ajec.43.4.03]
41. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. Implementation Sci 2013 Dec 1;8:139. [doi: 10.1186/1748-5908-8-139] [Medline: 24289295]
42. Seward K, Wolfenden L, Finch M, Wiggers J, Wyse R, Jones J, et al. Improving the implementation of nutrition guidelines in childcare centres improves child dietary intake: findings of a randomised trial of an implementation intervention. Public Health Nutr 2018 Feb;21(3):607-617. [doi: 10.1017/S1368980017003366] [Medline: 29173218]
43. Nathan N, Wiggers J, Bauman AE, Rissel C, Searles A, Reeves P, et al. A cluster randomised controlled trial of an intervention to increase the implementation of school physical activity policies and guidelines: study protocol for the physically active children in education (PACE) study. BMC Public Health 2019 Feb 11;19(1):170 [FREE Full text] [doi: 10.1186/s12889-019-6492-z] [Medline: 30760243]
44. Nathan N, Sutherland R, Hope K, McCarthy NJ, Pettett M, Elton B, et al. Implementation of a school physical activity policy improves student physical activity levels: outcomes of a cluster-randomized controlled trial. J Phys Act Health 2020 Sep 12;17(10):1009-1018. [doi: 10.1123/jpah.2019-0595] [Medline: 32919383]
45. Bélanger M, Humbert L, Vatanparast H, Ward S, Muhajarine N, Chow AF, et al. A multilevel intervention to increase physical activity and improve healthy eating and physical literacy among young children (ages 3-5) attending early childcare centres: the Healthy Start-Départ Santé cluster randomised controlled trial study protocol. BMC Public Health 2016 Apr 12;16:313 [FREE Full text] [doi: 10.1186/s12889-016-2973-5] [Medline: 27068684]
46. Adamo K, Wilson S, Harvey A, Grattan KP, Naylor PJ, Temple VA, et al. Does intervening in childcare settings impact fundamental movement skill development? Med Sci Sports Exerc 2016 May;48(5):926-932. [doi: 10.1249/MSS.0000000000000838] [Medline: 26656776]
47. 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;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
48. Welcome to Google Analytics. Analytics. URL: https://analytics.google.com/analytics/web/provision/#/provision [accessed 2021-11-28]
49. Yoong S, Grady A, Wiggers J, Stacey FG, Rissel C, Flood V, et al. Child-level evaluation of a web-based intervention to improve dietary guideline implementation in childcare centers: a cluster-randomized controlled trial. Am J Clin Nutr 2020 Apr 01;111(4):854-863 [FREE Full text] [doi: 10.1093/ajcn/nqaa025] [Medline: 32091593]
50. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
51. Jones J, Wyse R, Finch M, Lecathelinais C, Wiggers J, Marshall J, et al. Effectiveness of an intervention to facilitate the implementation of healthy eating and physical activity policies and practices in childcare services: a randomised controlled trial. Implement Sci 2015 Oct 25;10:147 [FREE Full text] [doi: 10.1186/s13012-015-0340-z] [Medline: 26498746]
52. Children's Services Award 2010. Fair Work OMBUDSMAN. URL: https://awardviewer.fwo.gov.au/award/show/MA000120 [accessed 2021-11-28]
53. Ward DS, Mazzucca S, McWilliams C, Hales D. Use of the Environment and Policy Evaluation and Observation as a Self-Report Instrument (EPAO-SR) to measure nutrition and physical activity environments in child care settings: validity and reliability evidence. Int J Behav Nutr Phys Act 2015 Sep 26;12:124 [FREE Full text] [doi: 10.1186/s12966-015-0287-0] [Medline: 26410387]
54. Dodds P, Wyse R, Jones J, Wolfenden L, Lecathelinais C, Williams A, et al. Validity of a measure to assess healthy eating and physical activity policies and practices in Australian childcare services. BMC Public Health 2014 Jun 09;14:572 [FREE Full text] [doi: 10.1186/1471-2458-14-572] [Medline: 24909075]
55. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
56. StataCorp. STATA. 4905 Lakeway Drive, College Station, TX 77845 USA 2020:2020.
57. The Australian Statistical Geography Standard (ASGS) Remoteness Structure. Australian Bureau of Statistics. 2012. URL: https://www.nintione.com.au/resources/rao/remoteness-structure/ [accessed 2021-11-28]
58. Dev DA, Byrd-Williams C, Ramsay S, McBride B, Srivastava D, Murriel A, et al. Engaging parents to promote children's nutrition and health. Am J Health Promot 2017 Mar;31(2):153-162. [doi: 10.1177/0890117116685426] [Medline: 28423928]
59. van de Kolk I, Goossens A, Gerards S, Kremers S, Manders R, Gubbels J. Healthy nutrition and physical activity in childcare: views from childcare managers, childcare workers and parents on influential factors. Int J Environ Res Public Health 2018 Dec 19;15(12):2909 [FREE Full text] [doi: 10.3390/ijerph15122909] [Medline: 30572600]

60. Finch M, Wolfenden L, Morgan PJ, Freund M, Jones J, Wiggers J. A cluster randomized trial of a multi-level intervention, delivered by service staff, to increase physical activity of children attending center-based childcare. Prev Med 2014 Jan;58:9-16. [doi: 10.1016/j.ypmed.2013.10.004] [Medline: 24145204]
61. Hardy LL, King L, Kelly B, Farrell L, Howlett S. Munch and Move: evaluation of a preschool healthy eating and movement skill program. Int J Behav Nutr Phys Act 2010 Nov 03;7:80 [FREE Full text] [doi: 10.1186/1479-5868-7-80] [Medline: 21047434]
62. Langford R, Bonell C, Jones H, Pouliou T, Murphy SM, Waters E, et al. The WHO Health Promoting School framework for improving the health and well-being of students and their academic achievement. Cochrane Database Syst Rev 2014 Apr 16(4):CD008958. [doi: 10.1002/14651858.CD008958.pub2] [Medline: 24737131]
63. National quality standard. Australian Children's Education & Care Quality Authority. URL: https://www.acecqa.gov.au/nqf/national-quality-standard [accessed 2021-11-28]
64. Finch M, Wolfenden L, Falkiner M, Edenden D, Pond N, Hardy LL, et al. Impact of a population based intervention to increase the adoption of multiple physical activity practices in centre based childcare services: a quasi experimental, effectiveness study. Int J Behav Nutr Phys Act 2012 Aug 29;9:101 [FREE Full text] [doi: 10.1186/1479-5868-9-101] [Medline: 22929434]
65. Luecking CT, Dobson P, Ward DS. Barriers and facilitators of parent engagement with health promotion in child care: a mixed-methods evaluation. Health Educ Behav 2020 Dec;47(6):914-926. [doi: 10.1177/1090198120952040] [Medline: 32815417]

Abbreviations

BCT: behavior change technique
BCW: Behavior Change Wheel
EATS: Electronic Assessment Tool and Support
ECEC: early childhood education and care
Go-NAPSACC: Go-Nutrition and Physical Activity Self-Assessment for Child Care
HNE: Hunter New England
HPO: health promotion officer
MoU: memorandum of understanding
NSW: New South Wales
RCT: randomized controlled trial
SES: socioeconomic status

Edited by G Eysenbach; submitted 03.12.20; peer-reviewed by C Saldaña, H Belli; comments to author 14.02.21; revised version received 08.03.21; accepted 05.11.21; published 15.12.21.

Please cite as:
Barnes C, Yoong SL, Nathan N, Wolfenden L, Wedesweiler T, Kerr J, Ward DS, Grady A
Feasibility of a Web-Based Implementation Intervention to Improve Child Dietary Intake in Early Childhood Education and Care: Pilot Randomized Controlled Trial
J Med Internet Res 2021;23(12):e25902
URL: https://www.jmir.org/2021/12/e25902
doi:10.2196/25902
PMID:34914617

Original Paper

A Digital Platform to Support Self-management of Multiple Chronic Conditions (ProACT): Findings in Relation to Engagement During a One-Year Proof-of-Concept Trial

Julie Doyle[1], BSc, PhD; Emma Murphy[2,3], BA, MSc, PhD; Shane Gavin[1], BSc; Alessandra Pascale[4], BSc, MSc, PhD; Stéphane Deparis[4], BSc, MSc, PhD; Pierpaolo Tommasi[4], BSc, MSc; Suzanne Smith[1], BSocSci, MSc; Caoimhe Hannigan[2,5], BA, MSc, PhD; Myriam Sillevis Smitt[6], BSc, MSc; Cora van Leeuwen[6], MA, MSc; Julia Lastra[7], BSc, MSc; Mary Galvin[2], BA, HDip, PhD; Patricia McAleer[1], BSc, MSc; Lorraine Tompkins[2], BSc, MSc; An Jacobs[6], MSc, PhD; Marta M Marques[2,8], PhD; Jaime Medina Maestro[7], BSc, MSc; Gordon Boyle[1], BSc, MA; John Dinsmore[2], BSc, PhD

[1]NetwellCASALA, Dundalk Institute of Technology, Dundalk, Ireland

[2]Trinity Centre for Practice and Healthcare Innovation, School of Nursing and Midwifery, Trinity College Dublin, Dublin, Ireland

[3]TU Dublin, Dublin, Ireland

[4]IBM Research Europe, Dublin, Ireland

[5]Department of Psychology, National University of Ireland, Dublin, Dublin, Ireland

[6]Imec-VUB-SMIT, Brussels, Belgium

[7]Tree Technology, Madrid, Spain

[8]ADAPT SFI Research Centre, Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Julie Doyle, BSc, PhD
NetwellCASALA
Dundalk Institute of Technology
Dublin Road
Dundalk, A91 K584
Ireland
Phone: 353 429370497
Email: julie.doyle@dkit.ie

Abstract

Background: Populations globally are ageing, resulting in higher incidence rates of chronic diseases. Digital health platforms, designed to support those with chronic conditions to self-manage at home, offer a promising solution to help people monitor their conditions and lifestyle, maintain good health, and reduce unscheduled clinical visits. However, despite high prevalence rates of multimorbidity or multiple chronic conditions, most platforms tend to focus on a single disease. A further challenge is that despite the importance of users actively engaging with such systems, little research has explored engagement.

Objective: The objectives of this study are to design and develop a digital health platform, ProACT, for facilitating older adults self-managing multimorbidity, with support from their care network, and evaluate end user engagement and experiences with this platform through a 12-month trial.

Methods: The ProACT digital health platform is presented in this paper. The platform was evaluated in a year-long proof-of-concept action research trial with 120 older persons with multimorbidity in Ireland and Belgium. Alongside the technology, participants had access to a clinical triage service responding to symptom alerts and a technical helpdesk. Interactions with the platform during the trial were logged to determine engagement. Semistructured interviews were conducted with participants and analyzed using inductive thematic analysis, whereas usability and user burden were examined using validated questionnaires.

Results: This paper presents the ProACT platform and its components, along with findings on engagement with the platform and its usability. Of the 120 participants who participated, 24 (20%) withdrew before the end of the study, whereas 3 (2.5%) died. The remaining 93 participants actively used the platform until the end of the trial, on average, taking 2 or 3 health readings daily over the course of the trial in Ireland and Belgium, respectively. The participants reported ProACT to be usable and of low burden.

Findings from interviews revealed that participants experienced multiple benefits as a result of using ProACT, including improved self-management, health, and well-being and support from the triage service. For those who withdrew, barriers to engagement were poor health and frustration when technology, in particular sensing devices, did not work as expected.

Conclusions: This is the first study to present findings from a longitudinal study of older adults using digital health technology to self-manage multimorbidity. Our findings show that older adults sustained engagement with the technology and found it usable. Potential reasons for these results include a strong focus on user-centered design and engagement throughout the project lifecycle, resulting in a platform that meets user needs, as well as the integration of behavior change techniques and personal analytics into the platform. The provision of triage and technical support services alongside the platform during the trial were also important facilitators of engagement.

International Registered Report Identifier (IRRID): RR2-10.2196/22125

(J Med Internet Res 2021;23(12):e22672) doi:10.2196/22672

KEYWORDS

digital health; aging; multimorbidity; chronic disease; self-management; integrated care; longitudinal study; engagement; usability; mobile phone

Introduction

Background

Multimorbidity, the presence of 2 or more chronic conditions in an individual [1], is a major global health concern [2], and in many high-income countries, it is now considered the norm rather than the exception [3]. Some evidence suggests that it is more common in older adults, likely because of the aging population, whereas other research states that it is more prevalent in disadvantaged groups, such as those with lower socioeconomic status [3,4]. Prevalence rates suggest that, globally, 1 in 3 people live with multimorbidity, with rates of 65% in people aged >65 years and 85% in people aged >85 years and rising [5]. However, multimorbidity is also known to affect younger adults, with social deprivation being a key determinant in younger and middle-aged adults [3,6]. Multimorbidity reduces life expectancy, decreases quality of life and physical functioning, and negatively affects mental health [7]. It also results in health inequalities [4]. Some evidence suggests that reported outcomes, such as lower self-reported health, increased medication issues, and higher health care utilization, are poorer for older populations [8].

Self-management is recognized as an important component of care for those with multimorbidity to maintain good health [9,10]. Self-management can be defined as the actions taken by an individual to manage symptoms, treatment, emotions, and lifestyle changes as part of living with a chronic condition [11]. Compared with those with one chronic condition, people with multimorbidity experience more challenges with self-management, associated with numerous self-care tasks such as symptom monitoring, management of multiple medications, and liaising with multiple health care professionals [10,12-14]. For people with multimorbidity, self-management is an iterative process requiring constant readjusting and reframing to understand their conditions and the associated changeable symptoms to inform personalized self-management responses [15,16].

Numerous digital health technologies have been developed to support the self-management of single chronic diseases, primarily diabetes [17]. However, given the rise in the number of people managing multiple chronic conditions, it is imperative to consider how such technologies can be designed and implemented to deal with the additional complexities of multimorbidity, such as the management of multiple symptoms and self-care tasks. Prior work has noted that technologies that help those with multimorbidity manage their health will only be successful if they do not exert further burden or inconvenience on the user [18]. Furthermore, little attention has been directed toward the importance of understanding how to support people with multimorbidity through the use of digital interventions [19]. Integrating the management of multiple conditions onto a single platform, where users can monitor their symptoms and relevant lifestyle parameters (such as activity), interact with all their data, share their data, and receive educational support, could help to minimize the known burden of multimorbidity self-management. In recent years, a small number of researchers have begun to examine how to design digital health apps for multimorbidity self-management [20-27], including medication management [20,28], how to manage health care conflicts [21,24], and how those with multimorbidity collaborate and communicate with informal caregivers and health care professionals [25,27]. However, we are unaware of platforms that have been implemented to tackle multimorbidity or evaluated over longitudinal periods. Indeed, a recent systematic review has highlighted the lack of digital health platforms, that include architectures and analytics capable of fully supporting chronic disease self-management [29], beyond what is possible through simple *apps*. The review identified 7 papers, all of which support single disease management, and not multiple diseases.

The research on longitudinal engagement with digital health and wellness technologies, particularly for older adults, is lacking. Engagement is necessary to achieve intended and effective outcomes [30]. However, it is understood that the adoption of digital health technologies is low [18]. It has also been argued that older adults are not seen as the primary users of such technology [31] and are not ready to adopt it [32]. However, recent research goes some way to resolve this issue. A study examining the use patterns of 9051 users (mean age 50.4 years) with a diabetes management app over a period of 180 days found that older adults were more actively engaged

with the app than younger users [33]. Wei et al [34] studied habit formation using wearable activity trackers with 20 older adults who had been tracking activity for 6 months. They found a range of factors contributing to sustained use, such as contextual factors such as the placement of the tracker, how often it needs to be charged, and the presence of features such as goal setting and reminders. It is generally accepted that the successful design of a health and well-being intervention to motivate engagement requires a user-centered design process and an iterative approach to development [29,30]. This is particularly important for older adults, who may be unfamiliar with such technology and who may have additional interaction requirements because of the physical and cognitive effects of aging [31]. It is therefore crucial that designers of health technologies for older adults involve them closely in the design of technologies to ensure that they can benefit from them.

Our main objectives were as follows:

1. Design and develop a digital health platform (ProACT) to support people with multimorbidity to self-manage multiple conditions on a single platform, including monitoring a collection of symptoms and well-being parameters, helping users to understand relationships between their symptoms and conditions, providing education personalized to the person's condition profile, and supporting the sharing of data with a care network.
2. Evaluate ProACT in a year-long proof-of-concept (PoC) trial with people with multimorbidity supported by people in their care network to determine engagement, usability, and experiences with the digital self-management of a set of multiple symptoms and well-being parameters.

In this study, we present the ProACT platform that is designed to support people with multimorbidity to self-manage multiple conditions and results from a 12-month PoC trial of the platform with respect to engagement and usability. Our findings indicate that the vast majority of people with multimorbidity stayed engaged with the platform over the 12-month period, a novel finding with respect to this cohort. The participants found the platform to be usable and of low burden. Qualitative data from participant interviews indicated that participants experienced a number of benefits as a result of using ProACT, such as improved self-management, health, and well-being. For those who withdrew, barriers to engagement were poor health and frustration with the elements of the technology, particularly the sensing devices to monitor health parameters.

The ProACT Platform

The aim of the ProACT project was to design a single platform for supporting people with multiple conditions, specifically 2 or more of the following: chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), chronic heart disease (CHD), and diabetes, to self-manage. These conditions were chosen, as the World Health Organization identified the 4 key types of chronic diseases as cardiovascular, chronic respiratory diseases, diabetes, and cancer [35]. In addition, outcomes from a multimorbidity study by Barnett et al [6], with data collected from 1,751,841 people in 314 medical practices in Scotland, showed CHF, COPD, and diabetes as significantly linked disease conditions with related comorbidities (eg, hypertension). Although the platform was initially developed to support people self-managing combinations of these conditions, it is sufficiently flexible to allow new conditions to be added to the platform over time.

The ProACT platform was designed and developed using an extensive user-centered design process. This process involved interviews, focus groups, co-design sessions (hands-on design activities with participants; Figure 1), and usability testing before the platform's deployment in the trial. A total of 58 people with multimorbidity and 106 care network participants, including informal carers, formal carers, and health care professionals, across Ireland and Belgium participated in this process. The findings from the user-centered design process have been published elsewhere [21,28,36,37].

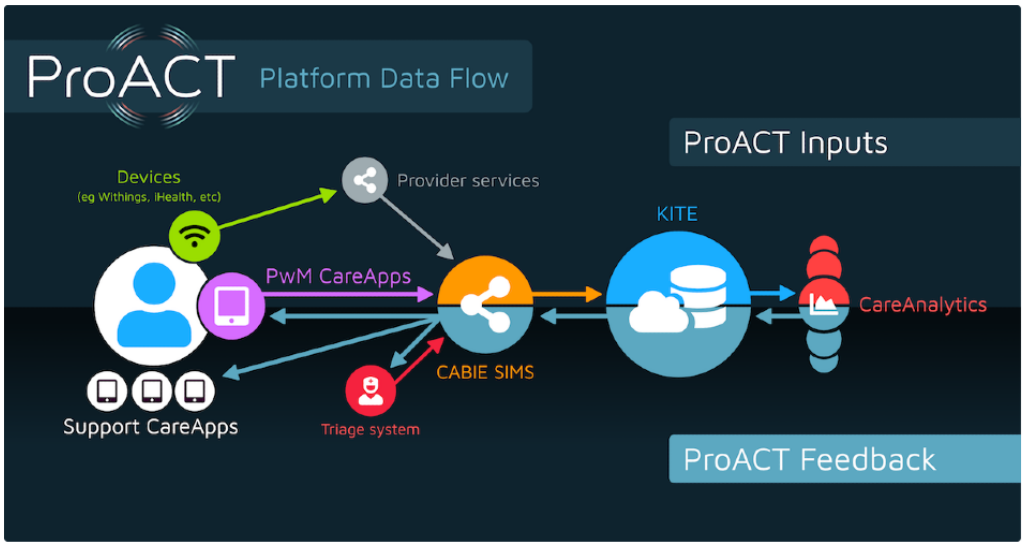
Figure 1. Participants in co-design workshops (A) choosing color schemes; (B) matching icons to labels.

The ProACT platform, designed and developed based on the findings from the above phases, has the following components:

- Measurement and sensing devices: off-the-shelf devices were used to collect health (blood pressure, heart rate, blood glucose, and pulse oximetry) and well-being (weight, activity, and sleep) readings from people with multimorbidity in their homes.
- ProACT person with multimorbidity CareApp: a web-based app delivered on an interactive device such as a tablet or smartphone that supports the monitoring of person with multimorbidity status and provides feedback and education to support health and wellness self-management. The CareApp has been designed to minimize the burden of self-management, for example, by helping people with multimorbidity prioritize condition management.
- ProACT care network CareApps: customized CareApp interfaces, developed for people in the person with multimorbidity's care network (informal carers, formal carers, and health care professionals), enables the viewing of the person with multimorbidity's health and well-being data, once the person with multimorbidity has agreed to share it.
- Context-aware brokering and inference engine (CABIE+): a source-agnostic data collection system to collect and organize sensor and self-reported data collected through the CareApp.
- Subject information management system (SIMS): a user management system allowed researchers or service organizations to personalize CareApps to individuals and manage, inspect, and analyze data.
- SIMS Triage: a version of the SIMS user management system was developed specifically for the clinical triage staff to view and respond to risk-stratified alerts from data collected by people with multimorbidity. This system provides clinical triage staff with a holistic view of the person with multimorbidity, displaying data relevant to their various conditions.
- KITE (Knowledge InTEgration) platform: a cloud-based infrastructure that allows the aggregation and processing of health data using a dynamic set of analytical components.
- CareAnalytics: innovative analytics to detect and react to data collected using ProACT. CareAnalytics support multimorbidity self-management, for example, by ensuring that any recommended content considers the person's overall condition profile, and behavior change, for example, by recommending educational content or highlighting a condition that needs attention, based on a person with multimorbidity's current health and well-being status.

The platform focused primarily on supporting the management of a collection of health symptoms and well-being parameters (relevant to the conditions of interest outlined above), helping users to understand the relationships between their symptoms and conditions, providing education relevant to individual conditions while also ensuring that this education considers multiple conditions, and provides a single app where people with multimorbidity or care network users could view all data and relevant content. The platform was developed to be compliant with the European Union's (EU's) General Data Protection Regulation (Regulation [EU] 2016/679) and underwent a data protection impact assessment before its deployment in the trial. Figure 2 illustrates the data flow between the different ProACT components. The following section provides a more in-depth overview of the components, primarily focusing on the person with multimorbidity CareApp.

Figure 2. Data flow within the ProACT platform. CABIE: context-aware brokering and inference engine; KITE: Knowledge InTEgration; PwM: person with multimorbidity; SIMS: subject information management system.



ProACT Person With Multimorbidity CareApp

Overview

In this section, we describe the CareApp used by the person with multimorbidity participants during the PoC trial. Further details on how we mapped specific user requirements to ProACT design features have been published elsewhere [37]. The following subsections outline the various features within the CareApp. The content of the CareApp is personalized at both *condition* and individual levels. For example, participants with diabetes and COPD will not see any features or content with respect to CHF management. Individuals with diabetes may

decide that they do not want to collect blood glucose data. This personalization is achieved within the SIMS system and is described below. The CareApp was developed as a responsive web app and is therefore accessible across a wide range of devices and platforms.

Symptom Monitoring and Reflection

Findings from our requirements gathering study highlighted that those managing multimorbidity need to monitor several health and well-being parameters on a regular basis; therefore, digital health apps to support multimorbidity need to consolidate multiple health and well-being parameters into a single app [21]. Access to data has also been shown to support behavioral change and the self-management of chronic conditions [16]. Where available, off-the-shelf sensors and devices were sourced to monitor these parameters, whereas questions delivered through CareApp's *Add Info* section supported the monitoring of additional parameters not measured through a device (eg, foot care for diabetes, fatigue, and breathlessness). Figures 3-5 show the person with multimorbidity CareApp. The flower of the dashboard (Figure 3) provides a quick overview of the person with multimorbidity's current status (eg, their current step count and their last blood pressure reading). Petals can be blue, orange, or pink. An orange petal highlights to the user that they have not taken a reading for a particular parameter in the past 5 days. Pink represents a *nudge* to the user to further explore the petal, for example if a reading is outside the person's defined *normal* threshold. A symptom reflection feature was designed as an extension of the flower dashboard design. Figure 4 shows what happens when a user clicks on the pink petal containing COPD symptoms in Figure 3. In addition to measuring and viewing their symptoms, this feature encourages users to *reflect* on their recent symptom readings with respect to their normal readings by asking them whether their reading is within a normal range for them. This feature was designed to support people in understanding their symptom readings rather than passively recording them. This may be particularly important when a person first starts self-managing with the platform, whereas later reflection may become more intrinsic.

Findings from our requirements study revealed that, sometimes, the management of one chronic condition can be forgotten, particularly if another is currently more acute [21]. The flower design, and the logic behind it, ensures that if a condition is not being monitored, it is brought to the attention of the person with multimorbidity. This could be a prompt or alert to monitor symptoms relating to that condition or educational content being pushed to them. The flower acts as a subtle, unobtrusive prompt—it is up to the person with multimorbidity to act on it. Our requirements study [21] and research by others [12,38] have identified that with multiple self-management tasks across different conditions, there is an increased need to support people with multimorbidity in prioritizing their activities, to reduce complexity and time burden. Therefore, within the CareApp, only the areas that require attention are highlighted. In addition to the flower petals, people with multimorbidity can view any of their historical readings through the *View Readings* section of the CareApp (Figure 5).

Figure 3. CareApp dashboard.

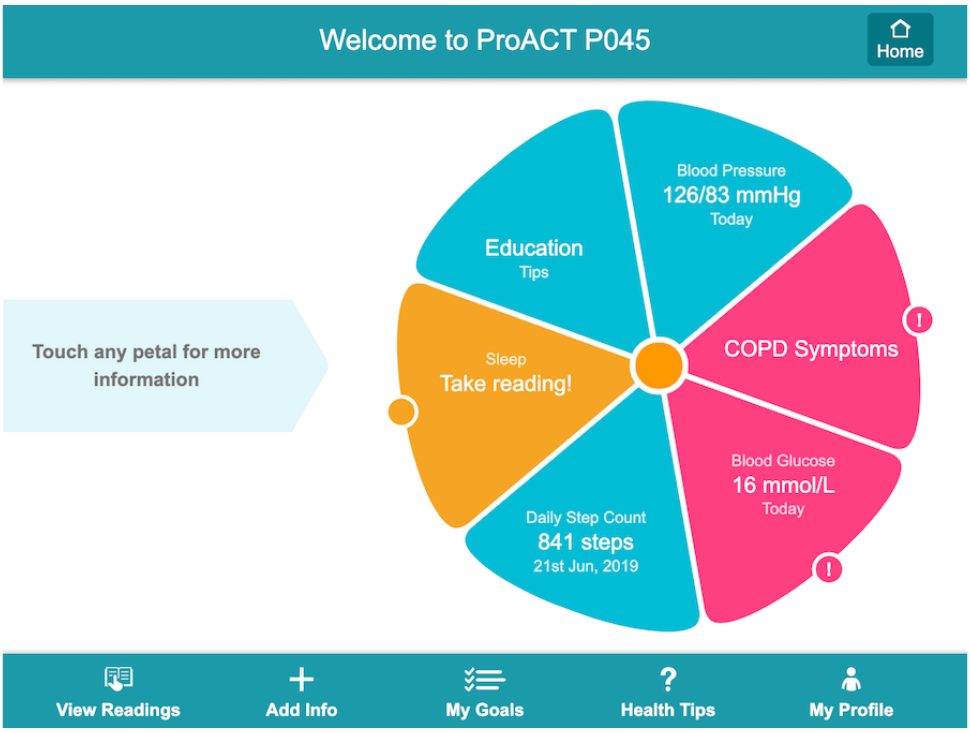


Figure 4. Reflection screen for abnormal chronic obstructive pulmonary disease symptoms.

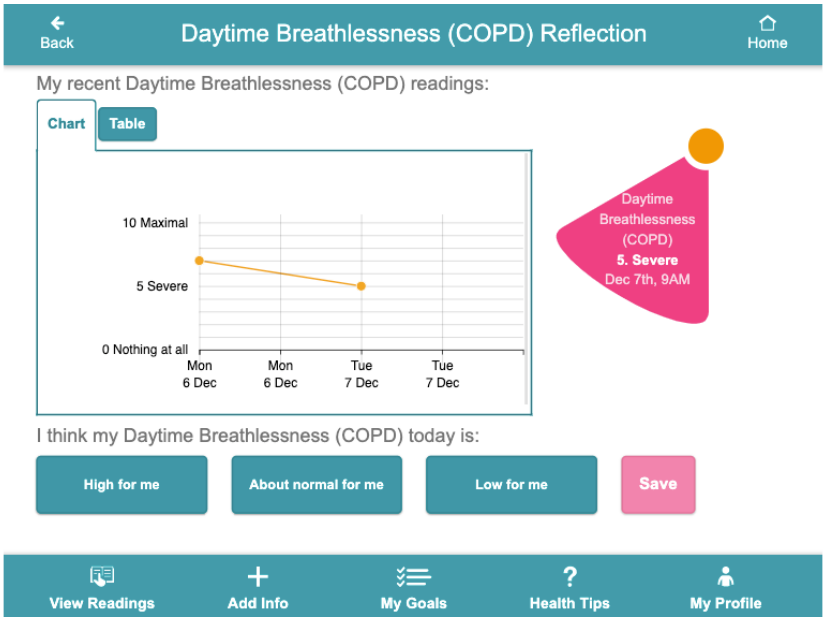
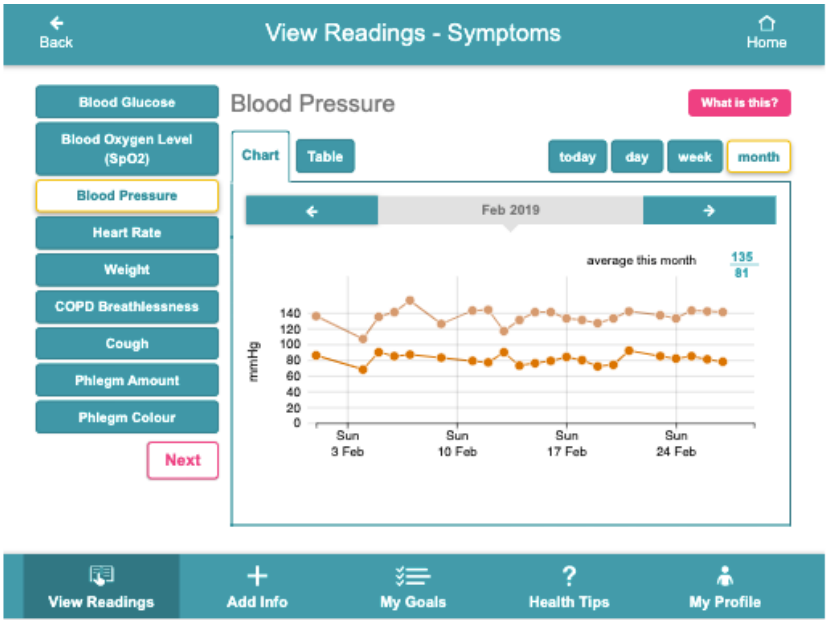


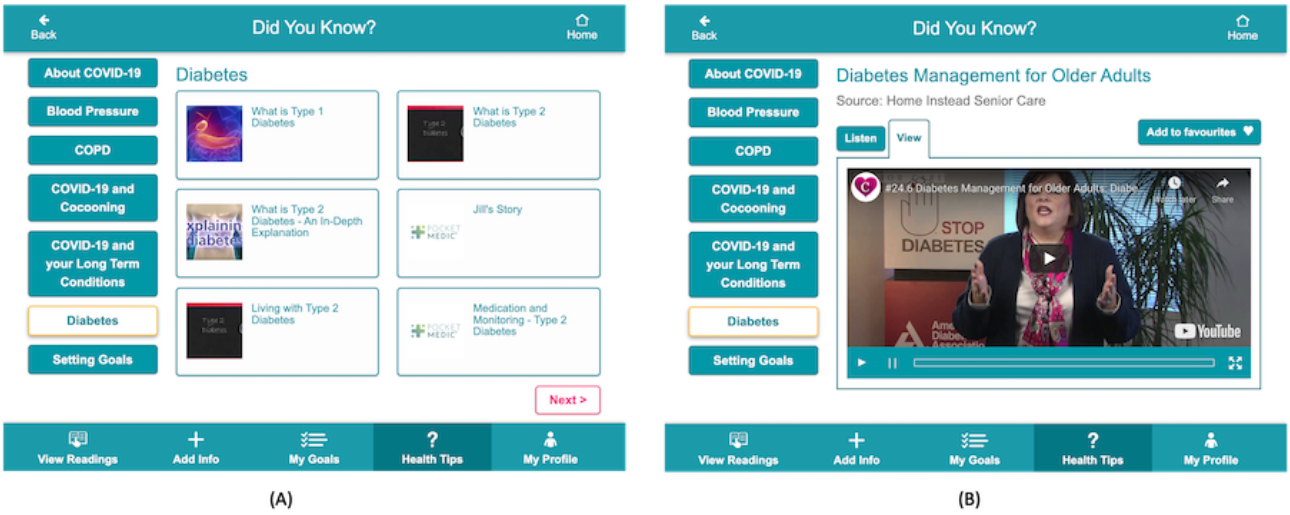
Figure 5. Data trends over time.



Education and Training

Findings from the requirements study, and those of others, highlighted that a lack of information is a significant barrier to both effective self-management and motivation to engage in self-management actions [21,39,40]. The selection of content for the education section of the CareApp and its planned delivery were therefore important tasks, and it was important to ensure the provision of trusted, reliable information tailored to a person with multimorbidity's specific conditions and management needs (including device and CareApp training). Within the *Health Tips* section of the CareApp, there are two categories of content: *Did you know?* contains educational content relevant to self-management of the person's conditions and well-being (Figure 6), and *How do I?* contains custom-made video training content on how to use the devices and CareApp. Educational information for each disease was sourced from reputable sources known to people with multimorbidity (eg, from national health services in each country). Where possible, content was delivered in three modalities (video, audio, and text) to cater to differences in learning styles and accessibility. An important consideration for the management of multiple conditions is that any education provided to the person with multimorbidity considers their current health status and their complete condition profile. This is achieved through ProACT CareAnalytics, as discussed in the section *CareAnalytics*.

Figure 6. Educational content in ProACT (A)The library of educational content; (B) A tip in video format.



Personal Goals

Setting goals and progressing toward goal achievement are key features for supporting the self-management of multimorbidity. People with multimorbidity can set physical activity goals in the ProACT CareApp. An overview of their weekly progress, measured through their activity watch, and an intuitive interface to help them set different goal modalities (eg, steps, distance, and time spent walking), are accessible through the *My Goals* section of the CareApp. Through messages and prompts, ProACT users are supported in setting achievable and incremental goals. The goal recommender described in the *CareAnalytics* section below suggests realistic activity goals based on the user's most recent activity data. As the user progresses and surpasses their targets, prompts suggest more challenging goals. Similarly, if the person with multimorbidity has difficulty reaching their target, they can say why this is by choosing a reason from a predefined list (eg, they were unwell this week), providing some context for why a particular goal was not achieved. Our requirements gathering identified additional features with respect to goal setting, such as collaborative goal setting with the person with multimorbidity's care network and feedback on goal progress [36], which will be integrated into future versions of the CareApp.

The Care Network

People with multimorbidity should be empowered to be contributors in the selection of their care network. Research has shown that the effective self-management of chronic disease does not occur in isolation but often involves, and is directly influenced by, support from informal carers and formal carers, as well as health care professionals during clinical visits [41,42]. Our findings from the requirements gathering also highlight, however, that the person with multimorbidity is often the coordinator of their own care, given the lack of integration among health care providers. Therefore, it is important that the person with multimorbidity can choose whom, within their care network, can support and contribute to their digital self-management. A feature within the *My Profile* section of the person with multimorbidity CareApp supports this, whereby the person with multimorbidity can add someone to their network and choose what data to share with them. CareApps for informal carers, formal carers, and health care professionals are accessible on their own devices (all CareApps are web based and responsive), allowing these care network stakeholders to see person with multimorbidity data that have been shared with them.

Usability and Accessibility

The usability and accessibility of the technology are crucial to ensure that users can easily interact with it. The CareApp was designed and developed to comply with relevant accessibility guidelines from tools, which are as follows:

- The Web Content Accessibility Guidelines 2.0 [43]: these guidelines state that (1) content must be perceivable; (2) interface components in the content must be operable; (3) content and controls must be understandable; and (4) content should be robust enough to work with current and future user agents (including assistive technologies).
- British Broadcasting Corporation mobile accessibility guidelines [44]: the British Broadcasting Corporation standards and guidelines for mobile accessibility are a set of technology-agnostic best practices for mobile web content and hybrid and native apps.
- Automated accessibility checkers and tools: although not as accurate as manual audits, automated accessibility checking tools such as the Web Content Accessibility Guidelines 2.0 offer better accuracy. A checker compliance tool (website shutting down in April 2021 [45]) was used. In addition, as color and contrast are important accessibility features for older users, tools were sourced that automatically check that colors have sufficient contrast within the interface (such as the web aim color contrast checker; [46]).

An accessibility and traceability spreadsheet was maintained detailing each guideline and was used to record accessibility issues and outline how they were resolved during the design of the CareApps. Accessibility guidelines should only form a part of an inclusive user-centered design process, as the most definitive test of accessibility can only be evaluated with end users [47]. End users, including people with multimorbidity and

care network stakeholders, were involved as co-designers throughout the entire iterative design and development process. Usability sessions were conducted with person with multimorbidity users from the early design phases, and throughout the various time points of the trial, allowing us to observe and gather feedback on usability and accessibility. The results of these evaluations informed interface updates, further enhancing the usability and accessibility of the app iteratively across the project.

Context-Aware Brokering and Inference Engine

CABIE+ is a source-agnostic cloud-based data aggregation platform. CABIE+ was the primary exchange mechanism used to connect the distinct technology components in ProACT. This component was used to centrally collect data from all connected devices used by people with multimorbidity, to normalize these data for storage, to make these data available to additional CABIE+ and ProACT components (SIMS and KITE), and to provide a configurable processing pipeline allowing incoming data to be inspected and reacted to in real time.

Subject Information Management System

SIMS is an administrative tool that facilitates the management of trial technologies, provides an abstraction layer for managing multiple CABIE+ instances and provides researchers or service organizations with a user-friendly, centralized service for monitoring and inspecting the various elements of the ProACT platform. SIMS also provides a user-facing application programming interface, which was used in the creation of the ProACT CareApps.

SIMS provides a number of features to facilitate the setup and management of trials and data. End users can be added to the system and their personalized profiles can be configured; for example, entering demographic information, configuring conditions being monitored within ProACT, and linking sensors and devices to their profile (Figure 7). Adding and scheduling educational content and self-report questions for delivery through the CareApp was also done via SIMS (Figure 8). For example, various categories of education and self-report questions can be created and delivered to all participants or to those with specific conditions. Furthermore, only participants configured in the system as having heart failure will see educational content and symptom questions related to that condition. Participant data can be queried via the *Inspect* feature (Figure 9), which shows the most recent inputs for each participant, including how long ago they logged in to the system (green indicates within the past day, red over a week ago). Each parameter measured by the participant can also be queried and viewed as a chart or table or exported for further analysis. SIMS also allows for personalized symptom thresholds to be set and generates alerts when symptom readings fall outside these thresholds (eg, alerting a triage nurse when a person's blood glucose level is outside their normal parameters).

The SIMS triage interface (Figure 10) presents clinical triage nurses with a list of alerts for individuals. Alerts are generated when thresholds for different parameters, defined within the system, are breached. For example, an alert for high blood glucose is a reading over 14 mmol/L (configurable per participant). Participants' placement on the list is prioritized by their alert status, that is, those with a *red alert* status appear first. A tag also appears alongside the alert, indicating whether it is new or under review. Within the dashboard, the nurse can also view recently resolved alerts. Nurses can query the person with multimorbidity's health and well-being data (as shown in Figure 9) to give them a holistic picture of the person with multimorbidity's status before calling them to discuss their alerts. Nurses can also create notes with respect to alerts, thereby allowing for a rich description of the context linked to alert readings.

Figure 7. Subject information management system dashboard, which allows for personalization and configuration of CareApps for person with multimorbidity and inspection of their data (participant names removed from image).

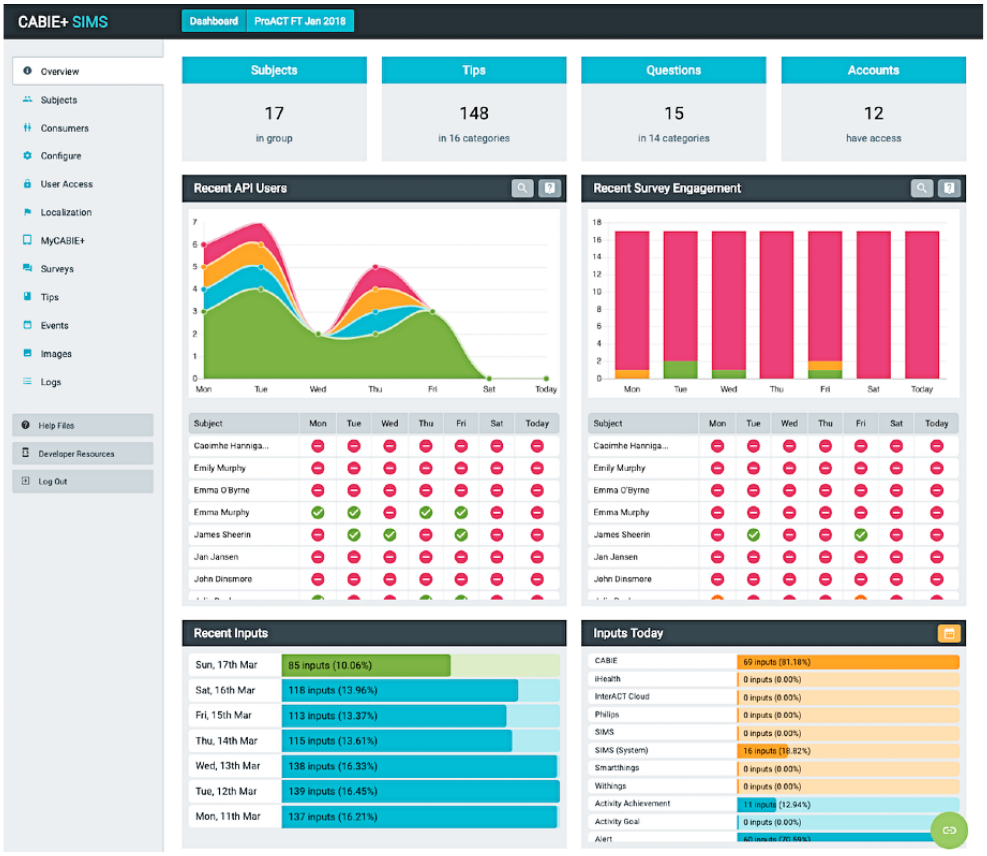


Figure 8. Subject information management system interface showing tip categories available and a number of tips related to heart failure.

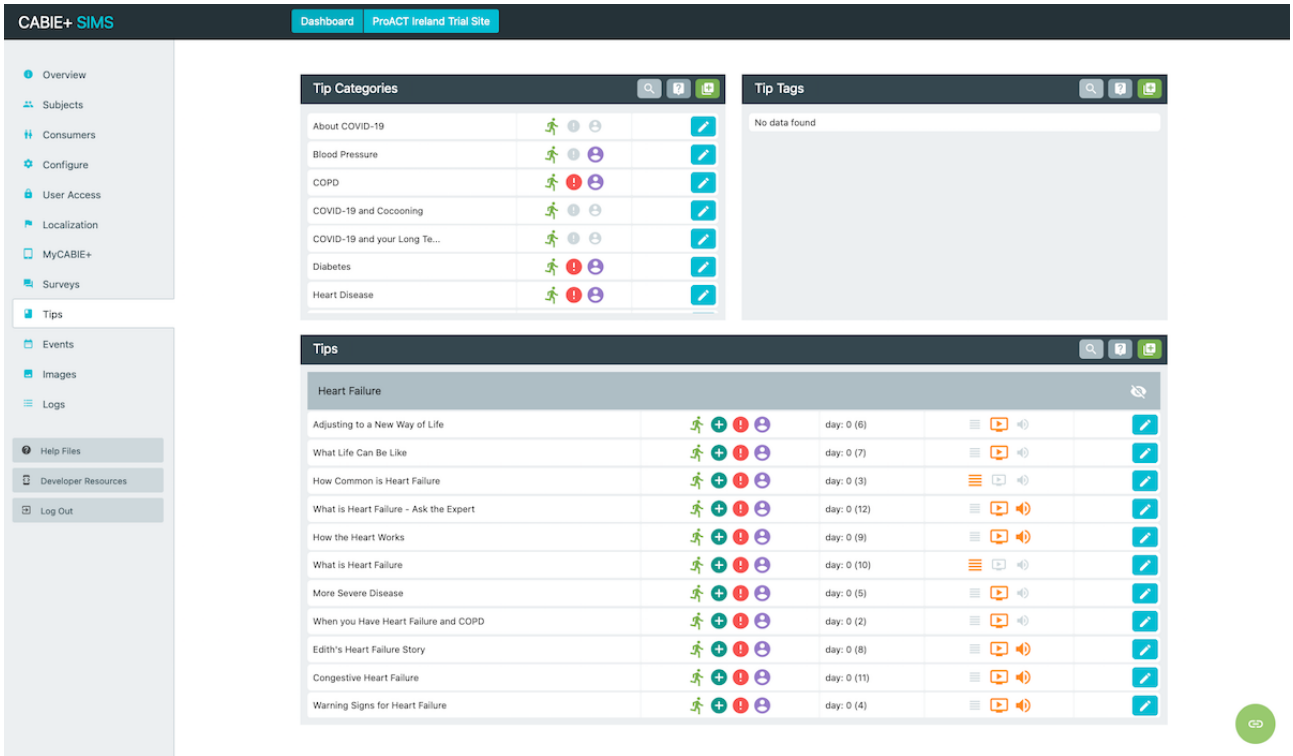


Figure 9. Inspect feature in subject information management system allows querying of data for each participant and shows their most recent inputs.

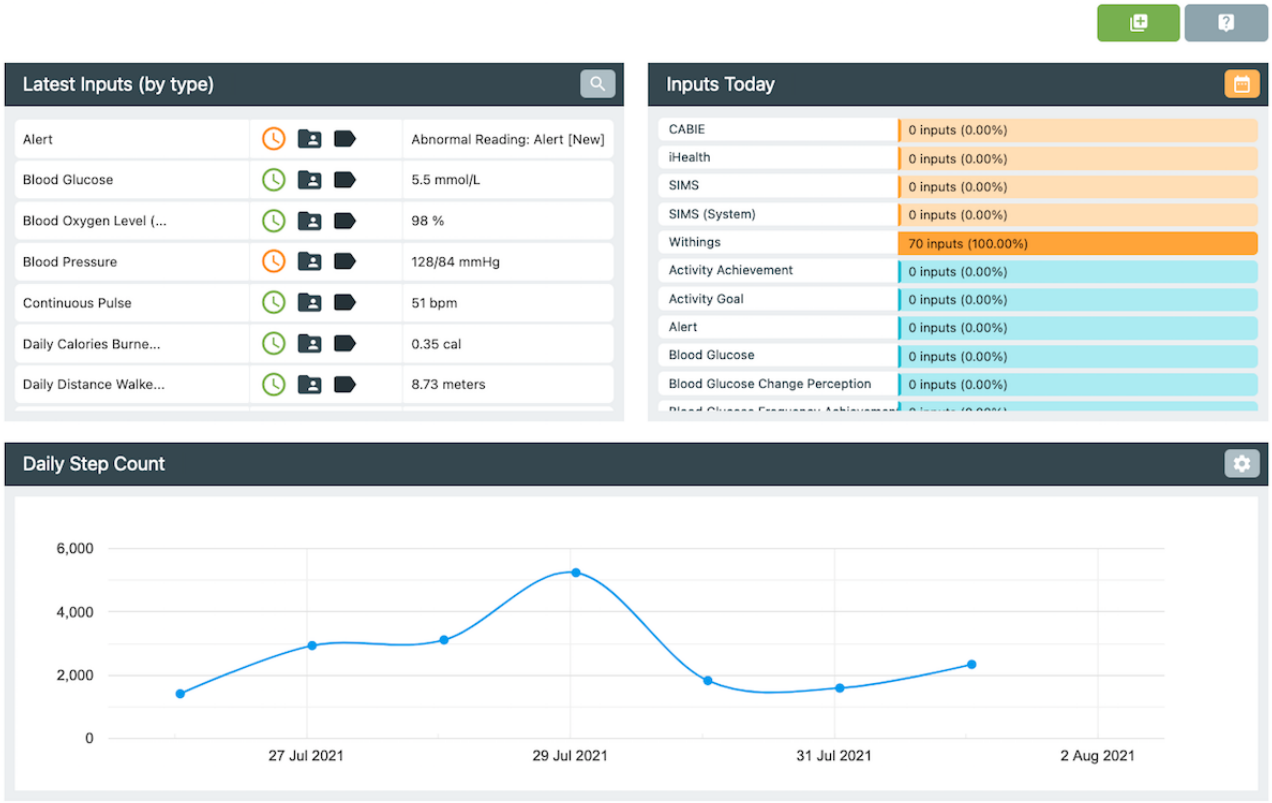
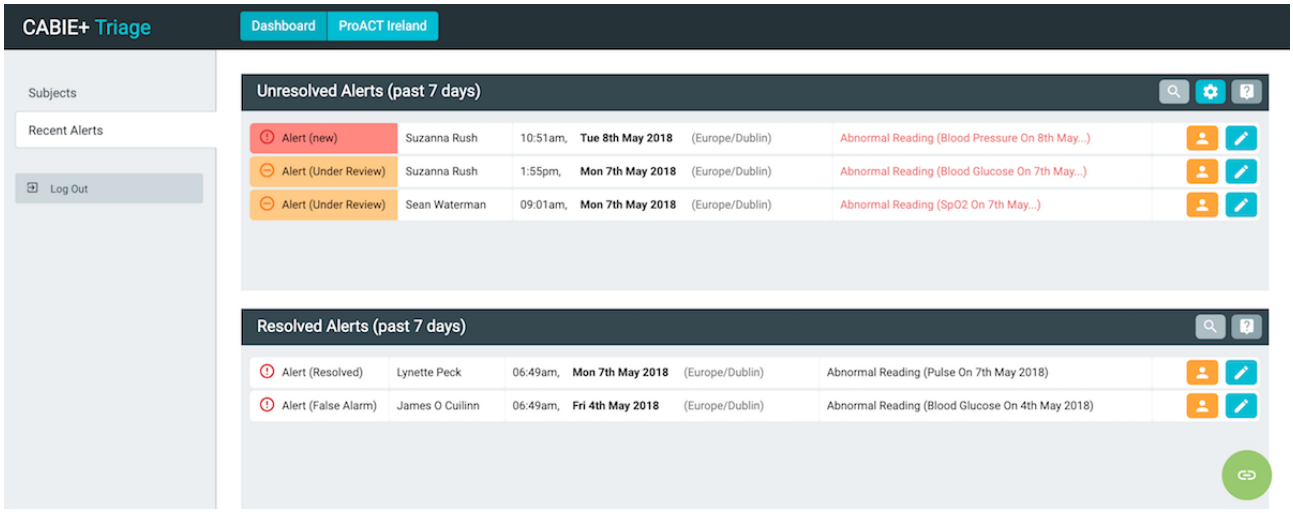


Figure 10. Subject information management system triage interface dashboard showing alerts (real names are not used).



Knowledge InTEgration

KITE is a cloud-based platform providing a reliable storage system for large volumes of data. In addition to being highly scalable, it eases data maintenance and access through an easily configurable data access configuration and transparently kept traceability metadata. KITE is exposed as a set of authenticated services for managing deidentified health data and coordinating collaboration among data providers, system data analytics, and data consumers. Figure 11 shows the actors of a typical workflow within KITE: providers push new deidentified data (eg, sensor data, self-reports, questionnaire results), which read the already deidentified data and create new results, whereas consumers read analytics' output and show it to the final users (people with multimorbidity and care network members). In a real-world scenario, analytics can selectively share output among themselves, allowing a complex CareAnalytic to be decomposed into a set of simpler and smaller analytics before reaching a data consumer.

From a logical point of view, KITE acts as a gate for all communications, modeled in an asynchronous fashion to ease maintenance and monitoring. Figure 12 shows the data flow diagram within ProACT (for simplification, CareAnalytics are represented with one box only; however, they might be implemented as a set of simpler subcomponents). The diagram is divided into three main areas: (1) data fetching, (2) personal information, and (3) deidentified information. The left section

lists all the data collected and pushed into the monitoring system managed by CABIE+. The data are deidentified and sent to the analytics that deal with deidentified information (right section). Although personal information data are stored on CABIE+, the deidentified information data (used by analytics) are stored in KITE.

Within KITE, CareAnalytics have two main purposes: making recommendations to people with multimorbidity (eg, by suggesting a goal or picking the most appropriate training material) and augmenting data (eg, creating aggregation or person-centered results).

Figure 11. High-level Knowledge InTEgration architecture. KITE: Knowledge InTEgration.

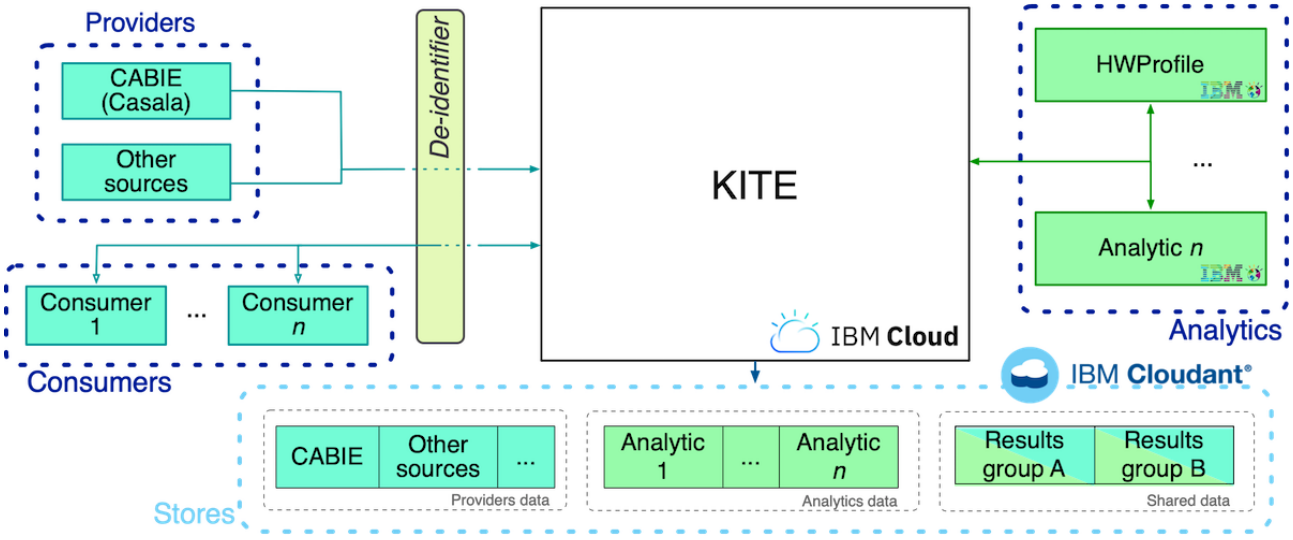
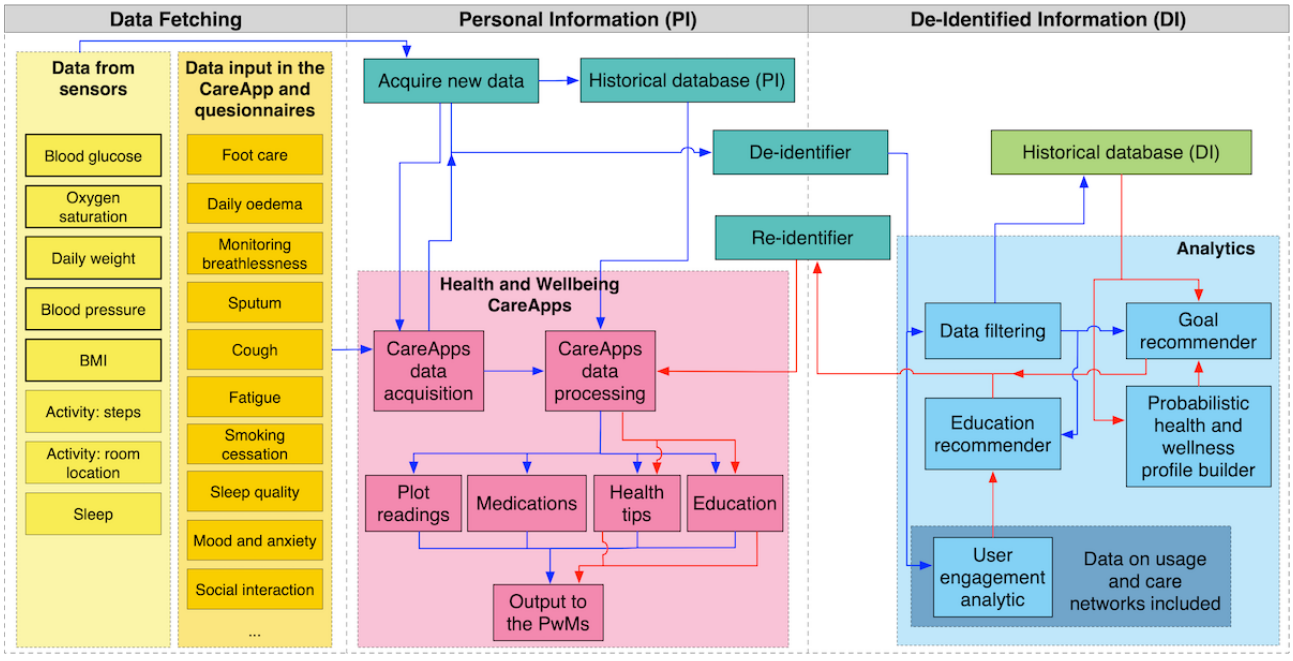


Figure 12. Data flow diagram implemented in ProACT. Blue arrows: data exchanges that occur with any new data input; red arrows: data exchanges happening at regular intervals, depending on the analytic. PwM: person with multimorbidity.



CareAnalytics

A *CareAnalytic* in ProACT is defined as a contextually aware procedure or algorithm that can detect and react to patterns in current or historic data available to ProACT systems. The current catalog of CareAnalytics in ProACT is as follows:

- *Data Cleaner*: this analytic classifies the different types of data in terms of quality (or reliability). In particular, the analytics that require vitals as input work on the results of the data cleaner instead of working on raw data. In addition, it calculates thresholds (ranges considered as normal) for the vitals that serve as an input for the *Education Recommender* (described below).
- *Goal Recommender*: this analytic supports people with multimorbidity in setting the weekly activity goals. The user can choose to set a goal in terms of distance, steps, or minutes spent walking. To suggest a goal value, the recommender considers the person with multimorbidity's physical activity over the previous week and whether they have met their previous goals. It also considers physical activity guidelines for older adults and adults with chronic conditions [48] to avoid giving major leaps in

recommendations. Finally, a maximum threshold (18,000 steps per week) was set as the total recommendation to ensure that the levels of activity were not too high for patients with chronic conditions and older adults [48].

- *Education Recommender*: this analytic facilitates the presentation of personalized educational and training materials within the CareApp. It uses thresholds on vitals; therefore, if any parameter has reached a value outside the normal range, then the person with multimorbidity will be presented with relevant educational material.
- *User Engagement Analyser*: this analytic tracks the person with multimorbidity's use of sensing devices and the ProACT CareApp.
- *Probabilistic Health and Wellness Profiler*: this is a probabilistic model of the person with multimorbidity describing them through several dimensions. Specifically, a categorical Bayesian network was learned using the open The Irish Longitudinal Study on Ageing (TILDA) data set. The TILDA data set holds data on 8504 individuals aged 50 years and above [49]. Variables were selected to cover several dimensions relevant to the conditions being monitored by ProACT and other relevant parameters such as vitals (eg, symptoms such as blood pressure), assessments (eg, nonsensor data such as fear of falling), self-reported wellness (eg, sleep quality), behaviors (eg, steps or physical activity), demographics and conditions (including the ProACT conditions COPD, CHF, CHD, and diabetes), and relevant comorbidities, such as additional chronic conditions, depression, and anxiety. The main goal of the model is to provide a comprehensive overview for a specific person with multimorbidity and their health and well-being states across different dimensions. Furthermore, the Bayesian network model could be used to infer the level of a missing variable based on the available measurements. For instance, a possible output would be the probability of a low physical activity score for a male aged 70-75 years with high blood pressure and diabetes. Some or all of these outputs can be used by other analytics. Further information can be found in Deparis et al [50].

During the ProACT PoC trial, detailed below, the *Data Cleaner*, *Goal Recommender*, and *User Engagement Analyser CareAnalytics* were implemented and used within the ProACT platform. The development of the *Health and Wellness Profiler* and the *Education Recommender* only occurred as data became available during the trial to help build and validate these analytics, and as such these were not present within the ProACT platform during the trial. However, the goal is to integrate these into future versions of the platform to further enhance personalization and multimorbidity management.

Methods

Inclusion Criteria and Recruitment

Trials occurred in Ireland and Belgium. The inclusion criteria for people with multimorbidity were that they were aged >65 years and had a diagnosis of two or more of the following diseases: diabetes, COPD, CHD, and CHF. It should be noted that 1 participant in Belgium was aged 60 years. This participant had initially recorded an incorrect date of birth on the screening documentation, and this was not discovered until after the participant had already begun the trial. This person was excluded from the core analysis presented in this paper. However, a separate analysis of their data was conducted, and the outcomes reflected those participants aged 65 years and above. Participants were recruited from a number of sources, including social groups for older adults, condition support groups, social media, radio and local newspaper advertising, formal care organizations, health care professionals, pharmacists, and living lab agencies. People with multimorbidity could also nominate up to 5 people to be part of their care network (including informal and formal carers and health care professionals). A total of 73 care network participants consented to participate. However, this paper focuses solely on the person with multimorbidity participants.

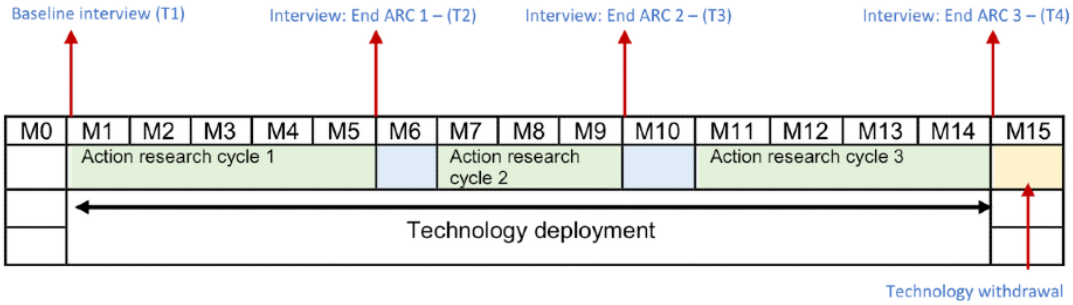
Study Design

The study was a PoC trial, which used an action research design, to allow for continuous feedback from participants and refinement of the ProACT platform throughout the trial. There were 3 action research cycles (ARCs) and 4 time points (T1-T4) of data collection, involving collection of questionnaire data, semistructured interviews, and usability testing (Figure 13). Refinements were made to the ProACT platform based on the feedback from participants at the end of each ARC. The same study design and methodologies were used across Ireland and Belgium to ensure consistency and comparability. The specific objectives were as follows:

- Evaluate the *usability and acceptability* of the ProACT CareApp and devices
- Evaluate *user adoption and satisfaction* with the technology and services
- Evaluate *experiences of using* ProACT

In this paper, we focus on the findings from studies that relate to engagement with the platform and usability.

Figure 13. Study timeline across action research cycles. ARC: action research cycle.



Procedures

Each participant was provided with an iPad, a smart watch to monitor sleep and activity, a digital weight scale, and a digital blood pressure cuff. Those with diabetes were provided with a blood glucose monitor (glucometer), whereas those with COPD were provided with a pulse oximeter to measure oxygen saturation levels. During the initial visit to the participant's home, the devices were set up by the researcher, and training was provided on using the devices. A second visit, approximately 1 week later, introduced the ProACT CareApp to the participants, and further training was provided. Each of the participants received a detailed paper-based training manual, while training videos on how to use the devices and CareApp were also available in the *Tips* section of the CareApp. The participants were asked to use the ProACT platform as they wished to evaluate how ProACT would fit into their lives naturally. During the trial, data from the participants were monitored by a clinical triage service in each country, appointed through a tendering process. The triage staff determined a protocol for alerts and escalation procedures in the case of an alert. The alerts were set within the SIMS triage platform described above. These were initially set at a global level (eg, the same blood glucose thresholds were set for all diabetic participants) and as the trial progressed, these could be individualized based on a person's normal range or input from their health care professional. The triage staff responded to alerts between 9 AM and 5 PM, Monday-Friday (hours were limited because of project budget) and made monthly check-in phone calls. Participants were made aware of the hours of triage, both through their project information sheet and a weekly pop-up message in the CareApp. Participants also had a helpdesk number they could call if they experienced any technical issues or wanted to request further training. A researcher was available to answer calls between 9 AM and 5 PM, Monday-Friday. Outside these hours, the participants could leave a message.

Data Collection and Analysis

The full study protocol, including details on all data collected during the trial and full analysis procedures, can be found in Dinsmore et al [51]. For the purposes of this study, data collection comprised interview data and questionnaire data on usability and user burden at each time point, which coincided with the end of each ARC. These data were collected at the participants' homes. Interview protocols at time point T1 covered motivations and expectations, whereas interviews at later time points covered a range of topics with respect to participants' experiences in terms of using the technology, benefits and challenges, their self-management practices, and care network support. Interviews were audio recorded and transcribed verbatim. A semantic thematic analysis [52] of these transcripts was then conducted using NVivo for Mac (version 11) in Ireland and MAXQDA in Belgium. A selection of transcripts was coded by 2 researchers to ensure a thorough iterative identification of a wide range of semantic themes. Initial broad coding was performed to identify the themes of interest, as covered within the interview protocols. Within these broader themes, an iterative thematic analysis was conducted to uncover the subthemes. Themes with respect to engagement are presented in this paper, whereas other themes (such as those relating to the self-management journey, experiences with the technology, behavioral change, and collaboration with the care network) are being submitted for publication elsewhere. Data on engagement with the system were logged through the ProACT platform, and metrics analyzed included the number of symptom readings per day and engagement with different sections of the CareApp.

Although additional data were collected during the trial, such as interview data with care network participants and triage staff, as well as the symptom and well-being data collected through the ProACT platform, the analysis of such data is outside the scope of this paper and will be published elsewhere.

Ethical Considerations

Ethical approval was received from 3 ethical committees in Ireland and 4 in Belgium. All procedures were in line with the General Data Protection Regulation for research projects, with the platform and trial methods and procedures undergoing data protection impact assessments in both countries. Written informed consent was obtained on an individual basis from participants in accordance with legal and ethical guidelines in each trial region, following a careful explanation of the study and provision of patient information and informed consent forms in plain language. All participants were informed of their right to withdraw from the study at any point without having to provide a reason for this.

Results

Overview

In total, 120 people with multimorbidity consented to participate, 60 in Ireland and 60 in Belgium. In Ireland, the average age of the participants was 74.23 (SD 6.4) years, and 60% (36/60) were male. In Belgium, the average age of the participants was 73.61 (SD 6.49) years, and the participants were predominantly male (43/60, 72%). Additional demographic data are presented in Table S1 of Multimedia Appendix 1. In this section, we present findings with respect to engagement with the ProACT platform, usability, and user burden and outline findings from the interviews with person with multimorbidity participants across the different time points that relate to engagement with ProACT. In particular, we focus on motivations to engage along with facilitators and barriers to engagement. At the end of the participant quotes, we identified the participant with the legend (ID, gender, age, inclusion conditions, time point, and country).

Engagement With ProACT

Over the course of the 12-month trial, the majority of participants remained engaged with ProACT. By the end of the trial, 3 participants had died in Ireland and 8 had withdrawn, resulting in 49 participants in Ireland completing the trial. In Belgium, 16 people withdrew, resulting in 44 participants completing the trial. Exit interviews were conducted with a subset of those who withdrew early and who consented to this. The reasons for withdrawal are discussed below in the *Barriers to Engagement* section.

Engagement with the sensor devices (used to record key symptom data) and ProACT CareApp were measured through

the platform. The heat maps in Figures 14 and 15 illustrate high levels of engagement with the ProACT devices for measuring key symptoms and well-being parameters, by the majority of participants in Belgium and Ireland, respectively. The horizontal bars of each graph depict one participant, with the start of the bar indicating their recruitment to the trial and the end of the bar indicating their exit from the trial. White bars indicate dropouts or those who died. During the trial, there was an average of 40 (SD 7.6) users in Ireland and 43 (SD 16.6) users in Belgium, taking measurements on a daily basis. The maximum number of participants recorded taking measurements on any day during the trial was 48 in Ireland and 60 in Belgium. Participants had an average of 2 daily readings in Ireland and three daily readings in Belgium.

Figure 16 shows how the ProACT CareApp was used across all trial participants: each row is a session and each cell is colored depending on the time spent on each section of the CareApp (the sections can be seen in the bottom menu in Figures 3-5). White cells are sessions with no visits to the correspondent section. With this in mind, *My Profile* and *Tips* were less visited sections. The rest of the sections are quite dense in the graph, which means that they have been frequently visited. More green cells are visible in *View Readings*, which means that participants spent more time checking their vitals than entering data or looking at their *Dashboard* (Figure 3). Furthermore, *Tips*, despite being a less visited section, has a significant number of green cells, which is expected as participants would likely spend more time here, viewing relevant educational videos or reading health tips.

Figure 14. User engagement in Belgium—daily symptom readings.

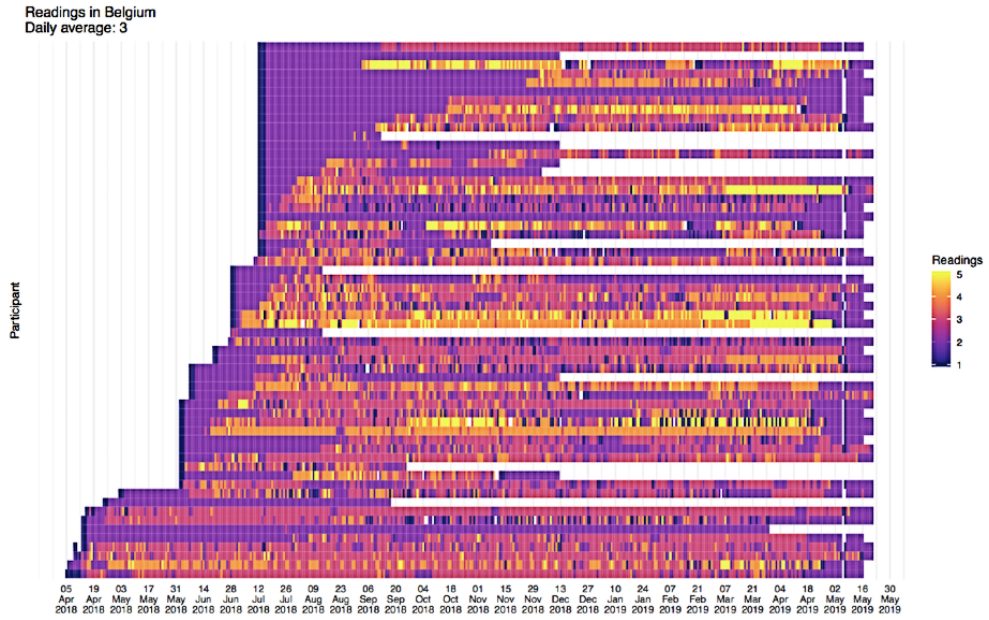


Figure 15. User engagement in Ireland—daily symptom readings.

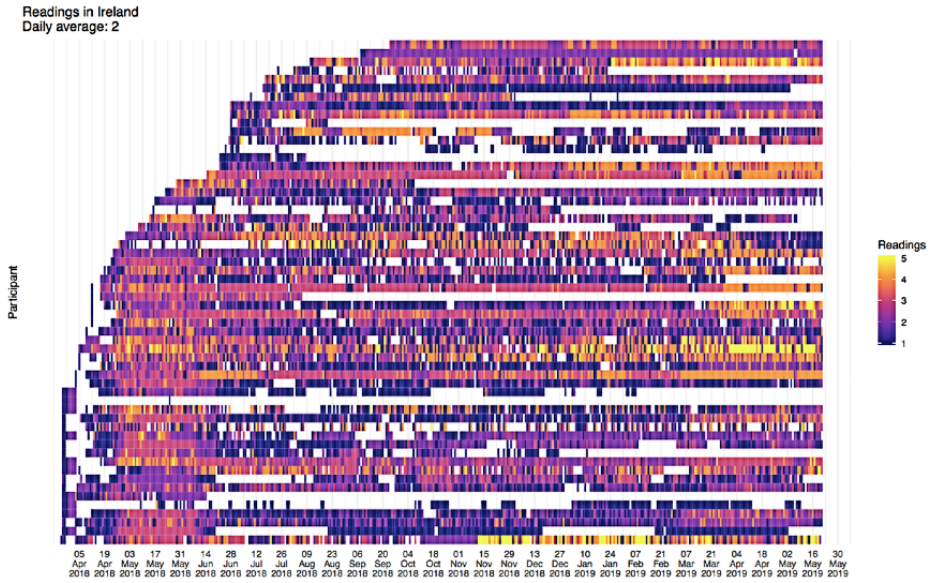


Figure 16. Heatmap of CareApp section visits across the trial period.

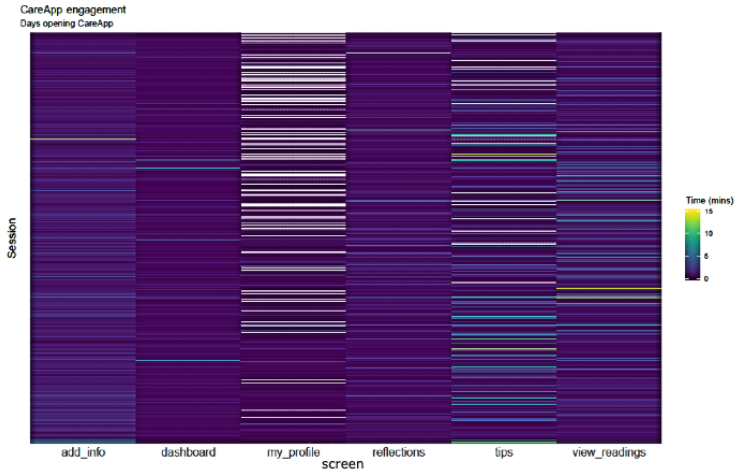
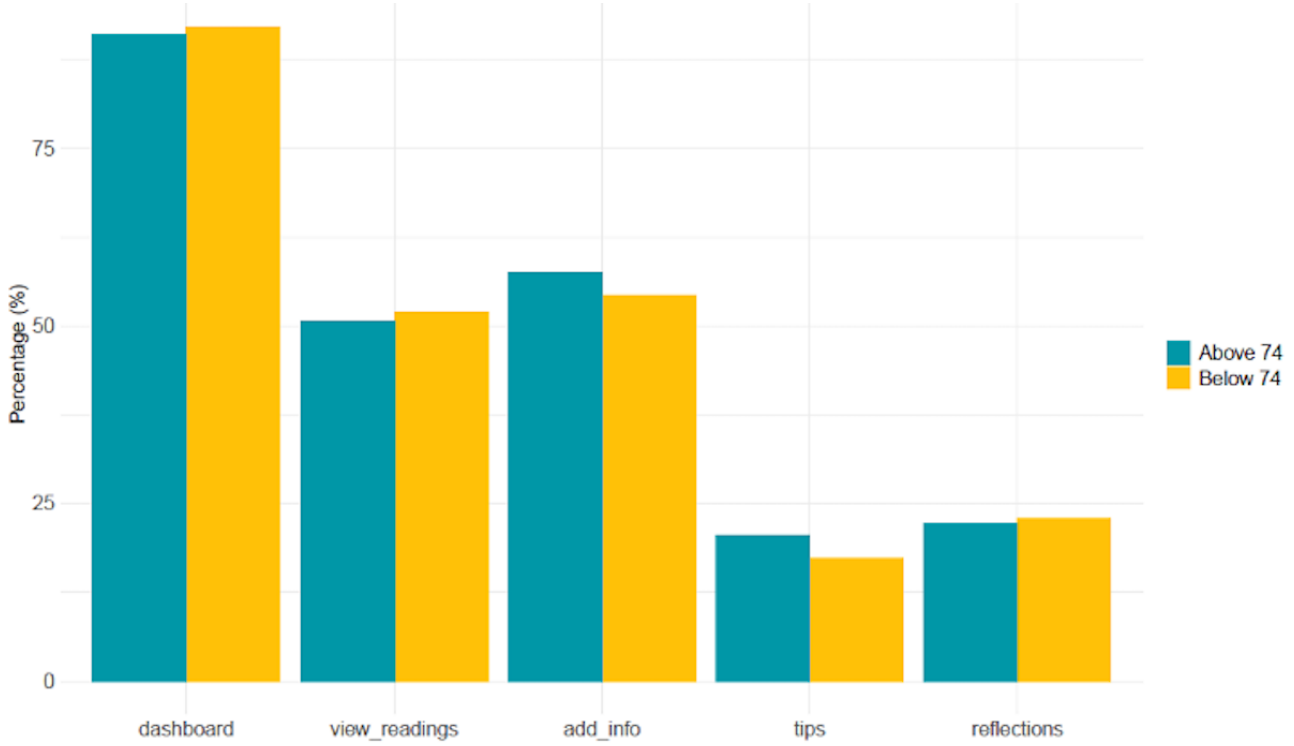


Figure 17 shows the percentage of time, on average, that trial participants spent in different sections of the CareApp. No significant difference was recorded for those aged less than 74 years compared with those aged over 74 years across the different sections of the CareApp. When a user opens the CareApp, they land on the *Dashboard*, which accounts for the highest percentage of section visits. The *View Readings* (where participants could see an overview of all their health and well-being data) and *Add Info* sections (where participants could answer the daily questions or add a manual symptom reading) were the most frequently visited sections by both age groups. Participants navigated to the *Tips* section less often. This may be because the *Tips* were used mostly at the beginning of the trial, as participants were learning about their conditions and self-management. Future work will examine this issue further. Similarly, the reflection feature (Figure 4), which asks participants if a particular daily reading is high or low for them, was not used as much as other sections of the CareApp, and our data show that its use declined over time.

Finally, the *Goal Recommender* was implemented within the platform for the final 8 weeks of the trial, primarily to test whether it worked as expected. Approximately half of the participants (53/96, 55%) set at least 1 activity goal during this period. Almost 75% (39/53) of participants set their goals in steps rather than meters or minutes. Furthermore, participants felt comfortable with the metric chosen—just 3 changed it over the course of the 8 weeks (from distance to steps).

Figure 17. Percentage of time participants engaged with different sections of the CareApp across the trial period, categorized by age group.



Usability and User Burden of ProACT

The participants completed the system usability scale (SUS) at T2, T3, and T4. The SUS is a 10-item questionnaire and is a quick and reliable tool for measuring the usability of both hardware and software products [53]. SUS scores range from 0 to 100. A value above 68 indicates that the system is usable. Table 1 indicates that SUS scores during the trial over time points T2, T3, and T4 remained relatively the same in Ireland and Belgium and above the threshold of 68, which indicates that participants considered ProACT to be usable. User burden was also measured at T2, T3, and T4 using the user burden questionnaire that measures the burden across six constructs, including difficulty of use, physical, time and social, mental and emotional, privacy, and financial [54]. No significant increases in burden were experienced over time.

Table 1. Usability and user burden scores across time points in Ireland and Belgium.

	Time points, mean (SD)		
	T2	T3	T4
Usability			
Ireland	74.6 (14.6)	72.4 (14.5)	75.6 (14.7)
Belgium	78.1 (17.0)	74.9 (16.8)	75.5 (20.8)
User burden			
Ireland	4.7 (4.0)	5.0 (4.4)	5.5 (4.6)
Belgium	3.2 (3.1)	5.2 (5.1)	5.8 (6.6)

Participant interviews provided insights into possible usability and burden issues associated with using the platform. The SUS did not distinguish between the ProACT CareApp and symptom-monitoring devices. However, participant feedback suggests that the majority of usability issues lay with using the devices, in particular the blood pressure monitor and the glucometer, which was a source of frustration for some participants:

> *I just stopped taking the blood pressure measurements. Because all the equipment wasn't working properly. In that it wasn't connecting to the iPad. There was one morning I timed it and it was only after spending half an hour connecting the thing on the screen.* [P009, male, 71 years, COPD+CHD+Diabetes, T3, Ireland]

Questions about accuracy were also noted: "Look last Friday it did not register sleep, why not?" [P106, male, 70 years, diabetes+CHD+CHF, T2, Belgium]. P103 (female, 79 years, diabetes+CHD+CHF, T2, Belgium) who had mobility issues stated that her limited steps do not always register.

Despite several reported device issues, participants felt that the platform was usable:

> *The app itself I think is very good. And I think it's helpful to get people to focus on a small number of things. That are key to their, you know their health. To also have a system that was easy to use, it is easy to use.* [P015, male, 82 years, diabetes+CHD, T3, Ireland]

> *Everything links in quickly enough...And I don't have a problem, there's no delay or whatever in it. It's straightforward. I think the equipment is excellent.* [P012, male, 67 years, diabetes+CHD, T3, Ireland]

Motivations to Engage

Overview

Two primary subthemes emerged with respect to motivation to engage at T1, which remained present at T2, T3, and T4 as reasons for continued engagement with ProACT:

- Improving self-management
- Improvement of health and well-being

Improving Self-management

At the start of the trial, improving self-management was noted by the majority of participants as the reason they wanted to engage with ProACT:

> *I haven't focused sharply enough [on health] and this is where I think this programme maybe will give me a bit of a kick in the backside, where it's very in your face and very visual and you know, just to see like say the blood pressure or the blood sugar readings.* [P033, male, 65 years, diabetes+CHD, T1, Ireland]

The participants felt that improved self-management would be achieved by gaining insights into their health from actively self-managing and seeing results, learning about conditions, learning what to avoid, and learning what to do if symptoms are changing.

At the start of the trial, a number of participants discussed how they hoped using ProACT would lead to increased confidence both in terms of managing their health and engaging with activity:

> *You know, seeing the results, it might give more confidence...I'd like to be able to have a round of golf and also I'd love to get back and play an odd set of tennis.* [P052, 78 years, male, diabetes+CHD, T1, Ireland]

P052, like many participants before beginning the trial, was fearful of undertaking physical activity in case it exacerbated

their conditions. Interview data from later time points in the trial show the participants became more empowered and confident in their self-management, as they used the ProACT CareApp to reflect on and compare data from a range of devices, in 1 place, across multiple symptom and well-being readings:

> *[The technology] gives you confidence, to be honest with you. And I quite like being able to take my own blood pressure and stats, so as I'm still in control.* [P031, male, 70 years, CHF+CHD, T2, Ireland]

P115 spoke of her fear before the trial, as an exacerbation could happen at any time:

> *I was living on a timebomb. Now, I do not live on any different timebomb than anyone else.* [P115, female, 73 years, CHF+CHD, T3, Belgium]

> *You are not as frightened because you know your readings and everything like that.* [P036, male, 72 years, COPD+CHD, T4, Ireland]

This was partly linked to higher levels of awareness:

> *Well, I suppose by being a bit more aware and a bit more in control, it's a positive thing and I suppose you have a bit more confidence as a consequence.* [P033, male, 65 years, diabetes+CHD, T2, Ireland]

Participants also reported making sense of how their actions affected their health, giving insight into what behaviors they could moderate or change to improve their various health readings:

> *Well it's telling me what's going on I suppose, I've a better idea myself and can relate the two together. If I haven't had activity, I can see how that is affecting blood sugars. I can see the relationship between the two, it's becoming clearer.* [P012, male, 67 years, diabetes+CHD, T3, Ireland]

Related to this, participants reported taking action based on their insights across their conditions and well-being, including changing their behaviors to improve their health readings. Participants also reported taking action with respect to their use of health care and interactions with health care professionals:

> *It is one of the reasons, through the use, that I went to my GP. That I said look, I am a bit worried. And then we looked for a solution and different medication and such. And that has changed due to [ProACT].* [P71, male, 69 years, COPD+CHD, T2, Belgium]

Notably, some participants reported going to the doctor less as they had more confidence from knowing their readings:

> *When I get an attack, the COPD or whatever flares, I can recognise [now] whether I need an antibiotic or not...Whereas before [using the technology], as soon as my breathing sort of laboured I'd be at the doctor.* [P045, 74 years, female, diabetes+COPD, T4, Ireland]

Many participants experienced benefits from such actions, which motivated engagement with the ProACT platform:

> *It's a whole health awareness...That's what [the technology] has created. Before this, I knew I needed to look after my health, but nobody does, and I got this system, now it's become a matter of routine.* [P053, male, 71 years, diabetes+CHD, T3, Ireland]

For P034, who reported struggling with mental health challenges, it was especially reassuring to see mood changes including periods of positive mood:

> *Oh, that was the highlight of my week, looking back on how I was doing...If I saw my graph was going up, and I wasn't as anxious yesterday as I was before, that would give me a bit of a boost.* [P034, 67 years, male, COPD+CHF, T4, Ireland]

Improving Health and Well-being

At T1, many participants expressed hope that engaging with ProACT and improving their self-management skills would in turn lead to improved health and a better quality of life, and this was the case for a number of participants:

> *My health has improved immensely...I've probably got the AFib [atrial fibrillation] under control...But, as I say, I've lost weight and I feel much better in myself. Mentally, physically, every way.* [P031, male, 70 years, CHF+CHD, T3, Ireland]

Indeed, throughout the trial, there were a number of critical clinical outcomes reported by participants in both Ireland and Belgium, where engagement with ProACT exerted a clear impact on a health outcome or directly influenced a change in treatment or care for the participant:

> *I've been in with my GP, he's increasing the medication. And we're still not there with it, so it may need to be tweaked further. So, without this system I might've been, you know, pretty oblivious to the fact that there is an issue there you know.* [P033, male, 65 years, diabetes+CHD, T3, Ireland]

Participants reported *symptoms stabilizing* as a result of implementing self-management practices:

> *I mean on average they [blood glucose readings] were 14 or 15 [14 mmol is threshold for a high alert], before this research started. And that was continuous. Now it's halved. Because I'm focused on improving it.* [P014, male, 70 years, diabetes+CHD, T3, Ireland]

Changes in food and diet were also widely reported, as well as increased activity levels, particularly as participants began to see the links between their activity and symptoms:

> *Well, the whole thing for me really is trying to get my weight down and keep my blood sugars down. And it's more stable...It's [ProACT] done everything really, it's gotten my exercise levels up because it focused me on it. It's got my blood sugars back in order.* [P048, male, 67 years, diabetes+CHD, T2, Ireland]

> *When I started [ProACT] up to now, I feel fitter...It is only 4000 steps, but I used to do maybe like 300 steps a day.* [P111, male, 76 years, diabetes+CHD+COPD, T3, Belgium]

Although participants in Belgium reported experiencing stabilization of symptoms and improvements in terms of activity and weight, few participants explicitly stated that they thought their health had improved. Most of the perceived health situations remained stable, except for those who endured health setbacks because of their illnesses or unfortunate incidents (eg, P117 had a fall and P81 had a car accident). However, 1 participant linked his engagement with the technology and the triage service with his continued ability to live alone:

> *They [family] were planning on doing that last year [sending him to a nursing home], now I do not have to, because I can make my own plan, because I can control myself.* [P74, male, 72 years, diabetes+CHD+CHF, T3, Belgium]

When participants attributed potential changes in health to using the platform, they were related to changes in attitude or behavior.

Potential Facilitators of Engagement

An important element of the ProACT platform, which was implemented during the trial, was the service provided alongside platform usage. A large theme that emerged across T2, T3, and T4 was *Service and Support*, which had 2 elements, namely, the triage service and the technical helpdesk. The triage service was one of the most commonly discussed topics by ProACT users in Ireland and Belgium**.** The majority of participants commented on the *reassurance* of having their readings monitored by the triage service. Most found it beneficial to know someone was looking out for them behind the scenes and making sure their readings were within a normal range:

> *The fact that...I knew someone was keeping an eye on it. And if that something was going out of kilter [range] that I'd get a nudge. And that's all I'd need...I know there's a nurse there who's looking at these figures.* [P015, male, 82 years, Diabetes+CHD, T3, Ireland]

Participants also noted how the triage nurses acted as a *backup* if they were to miss something, which gave further peace of mind:

> *It has given my wife a feeling of reassurance. People are keeping an eye on me. For me...it is also a reassurance.* [P81, male, 68 years, CHD+CHF, T4, Belgium]

As noted above, a number of participants reported clinical outcomes during the trial, and many of these were a result of triage intervention:

> *They [triage] had suggested maybe to go and get my cholesterol checked and my cholesterol was a bit high. It had been going a bit high and then my medication was changed, and it was as a result of ProACT.* [P039, female, 66 years, COPD+CHD, T4, Ireland]

> *[Triage nurse] told me get her into a hospital straight away otherwise you're going to have a body on your hands. Bring her out to the hospital and tell them the way she is. And I did that, and they kept her in...Yeah she [nurse] wanted to send an ambulance.* [Husband speaking during T3 interview for P018, female, 73 years, diabetes+CHD, T3, Ireland]

Social connection and personalized care were also identified as a subtheme within the triage service theme, whereby users felt that the triage nurses truly cared about their well-being. The majority of participants commented on how the nurses took the time to chat to them, enquire about their health, and found the phone calls to be a very pleasant experience: "What I think is so good about it, is that you have the feeling that someone thinks it is important what I do." [P101, male, 72 years, CHF+CHD, T3, Belgium]. Participants also appreciated the continuity of care and that the nurses always followed up on any issues, something many participants felt was lacking with their usual health care providers.

The research teams also operated a technical helpdesk that participants could call (Monday-Friday, 9 AM-5 PM) if they were experiencing any difficulties using the technology. When a participant called the helpdesk, the researcher attempted to troubleshoot with the participant over the phone. If this was not possible, then a researcher visited the participant to ensure that the issue was resolved. The responsiveness of this service was an important element in ensuring that participants could remain technically engaged with the ProACT platform and provide a fast and meaningful response to technical barriers encountered by participants, which might otherwise have resulted in disengagement:

> *I was helped along every time...[researchers] arrived straight away or as soon as possible to sort it out for me.* [P043, female, 77 years, COPD+CHF, T4, Ireland]

To offer context to the necessity of the helpdesk, there were a total of 355 calls to the helpdesk in Ireland over the course of the trial (helpdesk calls were not tracked in Belgium). Of these, 22% (78/355) were administrative calls, such as participants informing the team of an upcoming holiday or requesting clarification on interview dates. The remaining 78% (277/355) were related to tech support. Of these, 76 calls were requests for peripherals (eg, batteries for the blood pressure cuff and additional testing strips for the blood glucose monitor), 121 were for devices not working (eg, a blood pressure cuff that would not turn on), 50 related to data syncing issues (eg, if participants did not see their readings appear in the ProACT CareApp), 31 related to data accuracy concerns (eg, if a participant's own blood glucose monitor had a different reading to the digital one provided with ProACT), and 29 involved pairing or connection problems (eg, accidentally turning off Bluetooth).

Barriers to Engagement

Although the majority of participants learned to master the ProACT technology over time, technical issues were a barrier that ultimately led to some participants withdrawing from the trial. For P007 and P037 in Ireland, the reasons for withdrawal were *frustration with the devices*. Both found the devices and apps complicated to use: "It was just too hard to do because you couldn't in the first place take the blood pressure anyway because it wasn't working." [P037, female, 82 years, COPD+CHD, Ireland]. Although care was taken to address

these issues, for example, by replacing devices and providing further training, the initial experience for these participants resulted in a negative perception of the technology, which led to their disengagement. Related to this was a lack of trust in the readings. Some participants, including P93 and P105 in Belgium, found that the digital glucometer provided for the trial gave higher readings than their own personal glucometer, which caused them to distrust data from the trial device.

The second most common reason for withdrawing from the trial involved the complexity of the participants' conditions. For some, monitoring with ProACT felt like an added burden. Three participants in Ireland who withdrew did so as they felt they had enough to deal with in managing their conditions or with health complications which had arisen since the start of the trial. This was expressed by an informal carer for P007: "I think P007 has enough to contend with at the moment" [IC for P007, female, 78 years, diabetes+CHD, Ireland] and by P037 who felt under pressure to generate readings:

> *Well I was conscious of how much I was walking, you know and I'd walk up and down to the shed a whole lot of times if I wasn't out. So that was it, but I couldn't stick that for a year.* [P037, female, 82 years, COPD+CHD, IE]

P007's informal carer noted that readings generated concern: "Sometimes it was a little bit scary when you were getting those readings." P105 (76 years, female, diabetes+COPD+CHD, Belgium) stated that she experienced an extra burden by being confronted by some of the information provided by ProACT. She said that as a patient with diabetes, she already needed to be aware of a lot and that the pressure she felt from seeing the additional readings, such as activity, was experienced as too much of an extra burden.

Discussion

Principal Findings

This paper has presented the ProACT platform alongside findings with respect to engagement and usability from a 1-year PoC trial of the platform with 120 older people with multimorbidity across 2 EU countries. The findings demonstrate that the majority of participants actively engaged with the platform over the full duration of the trial, taking on average between 2 and 3 readings every day and engaging with self-reporting and their data in the CareApp. The qualitative data indicate that participants were motivated to engage as they found value and benefit in using the platform, including improving their self-management, gaining confidence, and experiencing health and well-being benefits. The results also indicate that participants found the platform usable and of low burden. Our study makes two primary contributions to the literature. First, we describe a novel, comprehensive digital platform, including CareApps, participant management, clinical triage, and analytics, and outline how it has been designed to support the self-management of multimorbidity. A recent systematic review highlighted that such platforms are lacking, and those that exist only target single disease management [29]. Second, we demonstrated that older adults were actively engaged with the platform over a 1-year period and we presented qualitative data to understand the possible reasons for this result. Textbox 1 summarizes the potential factors that may have enhanced engagement, which are elaborated upon in the remainder of this section.

Textbox 1. Likely impact of ProACT components on engagement.

Outcomes

- Engaging with symptom and well-being monitoring
- Engaging with reviewing data and acting on it

Potential Factors That Enhanced Engagement

- User-centered design (understanding needs and requirements)
- Usability of app
- All data on 1 platform
- Benefits experienced because of monitoring
- Technical support when needed, including face-to-face training during deployment and technical training via the application supplemented by paper-based materials
- Clinical oversight and support provided by triage nurses

Recently, researchers of human-computer interaction have begun to explore how technology can support the self-management of multimorbidity [18,21,22,24,55]. The majority of this research has focused on suggesting design requirements for technology, including understanding the requirements of people with multimorbidity and those who care for them [21-23,55], supporting communication about values with people with multimorbidity and health care professionals [27], managing health care conflicts [24], and medication management [20,28]. Research has also suggested potential reasons for those with multimorbidity not engaging with technology-supported self-management (including self-tracking being experienced as *work* and perceptions that health care professionals are not interested in self-tracked data) [18]. Our work in designing and developing the ProACT platform occurred before much of this recent research was published, with the trial beginning in April 2018. However, there are parallels in terms of design suggestions for multimorbidity management. Ancker et al [18] stated that

technologies for multimorbidity self-management will only be successful if they do not place further inconvenience or burden on the person. Indeed, earlier phases of our work, to understand user needs, highlighted how consuming it is living with and managing multimorbidity [21]. This was one of the primary drivers for designing a single platform for multiple self-management tasks, which includes features and analytics to help people with multimorbidity prioritize their conditions (eg, by highlighting on the dashboard *flower* those currently needing attention; Figure 3).

Caldeira et al [24] suggested three design guidelines for technology to support multimorbidity management: promoting conflict awareness (between disease and self-management approaches), supporting conflict resolution (to the most relevant approaches depending on symptom changes), and promoting patient expertise (to enact the approaches needed), which were considered while designing the ProACT platform. Considering patient expertise, the authors suggest that providing detailed health information and education can facilitate learning and contribute to expertise [24]. Our earlier findings also highlighted the lack of support people with multimorbidity receive in self-management [21], which identified the need for educational material, appropriate training on using self-management technology, and a triage service that would not only respond to symptom alerts but also motivate, support, and reinforce education presented through CareApp. People with multimorbidity in our study reported more awareness and increased confidence with respect to self-management, demonstrated the knowledge of their conditions, and what can impact them. With respect to designing for conflicts, CareAnalytics within ProACT have been designed to support the recognition and resolution of conflicts. For example, the Health and Wellness Profiler computes a probabilistic description of the person with multimorbidity, which is particularly important for multimorbidity, and all conditions can be considered while assessing the risk associated with the other variables in the model. For instance, the estimated probability of a given physical activity level is not the same if the person with multimorbidity has a fear of falling in addition to COPD. Using an automated goal recommender that considered individual patterns related to the performance of a given behavior (eg, goal achievement, optimal levels of activity), disease, symptoms, and usage of the recommendation feature, contribute to more personalized and adaptive digital health interventions. The goal recommender was only implemented for 8 weeks during the trial and was focused solely on physical activity goals. Future work with respect to this feature and analytic will focus on a more holistic approach to goal setting to better support the self-management of multimorbidity, in line with findings identified in our previous research [36]. Similarly, the education recommender will ensure that any recommended content considers the person with multimorbidity's condition and comorbidity profile as well as their current status to help them better understand potential conflicts and how to deal with them. With time, the artificial intelligence systems within ProACT are expected to learn from the data to provide more accurate recommendations and include additional parameters of personalization (eg, progress in other health behaviors, personalized education).

Despite this promising recent interest in understanding how to design technologies for the self-management of multimorbidity, there is little research on platforms that have actually been designed and implemented or evaluated for longitudinal periods. However, there is an abundance of research exploring how technology can support the management of single chronic diseases, including diabetes [56-58], COPD and related respiratory diseases [59], chronic kidney disease [60], and hypertension [61]. Our findings have many parallels with these studies. For example, Visser et al [58] explored the experiences of older adults using blood glucometers to monitor type 1 diabetes. The majority of participants reported that they sustained use of the device over time to know how much insulin was required before meals and because using the device gave a sense of safety and confidence. A greater confidence in practicing self-management was observed in our qualitative findings. The ability of the technology to provide novel insights into one's conditions and support the prevention of exacerbations (keeping symptoms under control) was also identified in our study, with similar benefits reported by Tenendez et al [59] in their study of older people self-managing chronic respiratory disease. There are also some differences between our findings and those of others. Tenendez et al found that participants took a reactive approach to self-managing, including the monitoring of symptoms, only doing so when they felt unwell [59]. The authors suggested that the potential reasons for this are that monitoring chronic respiratory conditions can be overwhelming and make people more conscious of their disease. They also suggest that long-term daily self-monitoring can lack value for those with chronic respiratory conditions, something that has also been highlighted by others [62,63]. Our findings contradict this result, thus indicating that people did find value and benefit in engaging in monitoring and seeing their data, with additional perceived value because of the triage service providing backup and peace of mind.

Our findings with respect to engagement with the sensor devices and CareApp indicate a high level of engagement with symptom and well-being monitoring and review. However, research has noted the importance of *effective* engagement rather than simply sustained or *more* engagement, with effective engagement being defined as *sufficient engagement with the digital intervention to achieve intended outcomes* [30]. Intended outcomes, from the perspective of digital health interventions, often relate to changes in behavior and improved health and well-being. For the ProACT trial, the key behavior under examination was engagement with the platform for self-management. Given that the ProACT platform represents a novel technology, the objectives of the trial were to evaluate usage, usability, and experiences to allow for the refinement of the technology before a trial to determine its effectiveness. This has been deemed a necessary approach to digital health intervention development and evaluation [64,65]. For example, Klasnja et al [64] suggest that evaluations of novel technologies should focus on human-computer interaction outcomes, including efficacy evaluations of specific intervention strategies such as self-monitoring and gaining a deep understanding of user experiences. Engagement with interventions has also been described as a precondition for effectiveness [30]. Although effectiveness was not specifically measured in our study, the

qualitative data indicated that participants changed their behaviors, including that they improved their self-management skills, got their symptoms under control, engaged in physical activity, experienced weight loss, and changed when and how they interacted with their health care professionals. Participants in Ireland also reported improvements in health. The analysis of the sensor and well-being data collected through the platform during the trial is currently underway, focusing particularly on symptom stabilization and physical activity to determine whether engagement with the intervention had a positive impact on these areas.

Previous research has indicated that the adoption of self-management technologies is limited [66]. Studies on the adoption of self-management technologies by older adults, particularly the oldest old, are lacking [67]. A number of studies have focused on conducting interviews, focus groups, or surveys with older adults to gauge interest in health technologies or barriers. Many of these reported negative attitudes. For example, Heart and Kalderon [32] found that older adults lacked interest in such technologies and perceived accessibility barriers. The authors concluded that such technologies would not be accepted by older adults unless they were very easy to use, useful, and there was support available to help with technical difficulties. The authors also suggested that when such factors are present, older adults can embrace self-management. On the basis of data from real-world engagement and participant feedback, our findings support this argument.

A number of possible reasons for sustained engagement during the ProACT trial have been identified. Throughout the trial, participants reported benefits, including improvements in health and well-being related to an increased ability to self-manage. Perceived benefit plays a role in technology acceptance and adoption [68,69] as well as in the activation of effective self-management behaviors [16]. Our findings are in line with those of a systematic review of studies examining older adults' usage of technology to support chronic disease management, which found that participants experienced greater self-efficacy because of technology-supported self-management, which enabled improved communication with health care providers, including personalized feedback and support [17].

Yardley et al [30] state that "successful intervention design demands a user-centered and iterative approach to development, using mixed methods and in-depth qualitative research to progressively refine the intervention to meet user requirements." A notable feature of the ProACT approach to designing and evaluating the platform was that it continually involved people with multimorbidity in a user-centered design process, including interviews, focus groups, co-design workshops, usability testing and evaluation activities, supporting regular, iterative updates to the ProACT platform. Our findings indicate that the majority of trial participants found the ProACT platform to be usable. Usability barriers are recognized within the literature as negatively impacting older adults' adoption of and engagement with technology [70,71]. Our findings also indicated that the participants found using this technology exerted low burden. This is an important finding given that there is significant work associated with self-management of chronic diseases, particularly multiple diseases and comorbidities [41]. Although participants had self-management activities outside of using the platform and their interactions with triage, such as attending appointments and managing medications, which were not evaluated as a part of the trial, it is promising that engaging in symptom monitoring, review, and education did not result in a high level of burden for the majority of participants. The usability and burden findings, along with the engagement data, indicate that the extensive user-centered design process during the project lifecycle cannot be underestimated; without this effort, it is possible that users would not have sustained their use of ProACT.

Research has noted that using digital health platforms for self-management without human support can negatively impact engagement, resulting in dropout and nonusage attrition [30,72]. Human services, including the triage and helpdesk, provided alongside the ProACT technology platform, appeared to be an important factor in terms of contributing to sustained engagement. Participants received feedback and advice on their health and well-being from the triage nurses and experienced continuity of care. Many clinical outcomes were also precipitated by the triage staff. Ancker et al [18] suggested that participants in their study lacked motivation to self-manage, as they perceived that health care professionals were not interested in their data. Furthermore, participants using ProACT had access to a technical helpdesk for the study duration if they experienced any technical issues. As is evident from the Irish helpdesk data outlined above, this was a necessary service because across the trial period for 60 participants, 355 calls were made to the helpdesk, 277 of which were related to technical issues, primarily related to the third-party devices provided to participants for symptom monitoring. Indeed, frustration with the sensing devices, including a poor usability, lack of reliability, and mistrust of the data they produced, were reasons for participants withdrawing from the trial. Given the importance of such devices in health self-management, device providers should strive to ensure that they are usable and reliable. Other research on the abandonment of health tracking devices found reasons that included technology being too complicated, too complex to learn, or failing to help people reach their goals [73].

Future research should also examine the potential role and necessity of helpdesk and triage services over longitudinal periods of time, as the needs of people with multimorbidity, as well as their self-management and technical skills change. Research should examine at what point or points of a self-management journey human support adds value to a digital health intervention. For example, some research has shown that support for self-management is mostly needed when a person first begins using a digital health intervention [56] to ensure adoption. However, research into the type and level of ongoing support that might be required for self-managing with multiple, complex chronic conditions is lacking. This also needs to be considered within the context of aging, as many older adults will experience a decline in health or the diagnosis of additional conditions and comorbidities over time. However, it is important to note that the levels of triage and technical support required also have implications in terms of costs of service provision; for example, the ratio of nurses to patients that is required to

provide an effective service. Hence, this aspect also needs to be further examined.

Limitations

There are several limitations to this study. First, given the resource limitations, the version of the ProACT platform described in this paper and evaluated in the trial did not integrate all of the features required to fully support the self-management of multimorbidity, with development initially focused on symptom and well-being monitoring, education, setting physical activity goals, and data sharing. The inclusion of features such as medication management, care planning, and goal setting beyond physical activity would be required to fully address the challenges of the self-management of multiple conditions. However, the addition of such features could also lead to an additional technological burden, which could negatively impact engagement. The integration of CareAnalytics not currently integrated within the platform and further development of all analytics could help minimize burden while also enhancing the platform's capability to handle recommendations that consider a person's full condition profile. For example, future work aims to further enhance the goal recommender in terms of personalization by considering the person's current health status, as indicated by the Health and Wellness Profiler analytic described above.

A further point with respect to the PoC study design (which was primarily to understand the feasibility of people with multimorbidity engaging with a digital health platform to self-manage their conditions) is that a control group was not included; therefore, it was not possible to assess the effectiveness of individual components of the ProACT platform and related services (access to triage and technical support). Future work will aim to examine the potential impact of these factors in more detail through controlled studies with people with multimorbidity. Implementing an iterative, user-centered PoC trial has helped justify the need for a large-scale controlled study. Therefore, we advocate that such an approach to designing and developing complex interventions for the management of multiple diseases should be considered before progressing toward large-scale studies.

Conclusions

This work is the first to present a digital health platform designed specifically to support multimorbidity self-management for older adults, with results showing high levels of engagement and retention over a 12-month period. There are several strengths to this work. We present a comprehensive digital platform that outlines how various components support the management of multiple diseases. There was active engagement from participants across all stages of the design, development, and evaluation of the platform. The trial involved a large number of participants across 2 EU countries over a longitudinal period. Few studies on digital health solutions to support the self-management of chronic conditions have focused on older adults. Neither do they have such large numbers (especially at the PoC level) nor do they examine long-term usage. Furthermore, comparatively little research has examined the design and evaluation of digital health platforms for people with multimorbidity.

Planned future work includes updating the platform to further enhance the management of multimorbidity and a pragmatic randomized controlled trial to evaluate the effectiveness (including potential cost-effectiveness) of the platform and triage service on quality of life, health care utilization, and a range of other measures, which will begin in 2022.

Acknowledgments

The ProACT project has received funding from the EU's Horizon 2020 Research and Innovation Programme under grant agreement 689996. MMM is funded by a Marie-Sklodowska-Curie fellowship (EU H2020 EDGE program grant agreement 713567). The authors would like to sincerely thank all the participants of this research for their valuable time. The authors wish to thank all consortium partners [74] for input across the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Person with multimorbidity demographics.
[DOCX File , 16 KB - jmir_v23i12e22672_app1.docx]

References

1. Boyd CM, Fortin M. Future of multimorbidity research: how should understanding of multimorbidity inform health system design? Public Health Rev 2010 Dec 10;32(2):451-474. [doi: 10.1007/BF03391611]
2. Hajat C, Kishore SP. The case for a global focus on multiple chronic conditions. BMJ Glob Health 2018 Jun 22;3(3):e000874. [doi: 10.1136/bmjgh-2018-000874]
3. Multimorbidity: a Priority for Global Health Research 2018. The Academy of Medical Sciences. URL: https://acmedsci.ac.uk/file-download/82222577 [accessed 2021-03-01]
4. Multimorbidity: Technical Series on Safer Primary Care. The World Health Organisation. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252275/9789241511650-eng.pdf?sequence=1&isAllowed=y [accessed 2020-03-01]

5. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev 2011 Sep;10(4):430-439. [doi: 10.1016/j.arr.2011.03.003] [Medline: 21402176]
6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2]
7. Azaïs B, Bowis J, Wismar M. Facing the challenge of multimorbidity. J Comorb 2016 Feb 17;6(1):1-3. [doi: 10.15256/joc.2016.6.71]
8. France EF, Wyke S, Gunn JM, Mair FS, McLean G, Mercer SW. Multimorbidity in primary care: a systematic review of prospective cohort studies. Br J Gen Pract 2012 Apr;62(597):e297-e307 [FREE Full text] [doi: 10.3399/bjgp12X636146] [Medline: 22520918]
9. Rijken J, Jones M, Heijmans M, Dixon A. Supporting self-management. In: Caring for People With Chronic Conditions. A Health System Perspective. London, UK: Open University Press; 2008.
10. Bayliss EA, Steiner JF, Fernald DH, Crane LA, Main DS. Descriptions of barriers to self-care by persons with comorbid chronic diseases. Ann Fam Med 2003 May 1;1(1):15-21 [FREE Full text] [doi: 10.1370/afm.4] [Medline: 15043175]
11. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [Medline: 12401421]
12. Eton DT, Ramalho DO, Egginton JS, Ridgeway JL, Odell L, May CR, et al. Building a measurement framework of burden of treatment in complex patients with chronic conditions: a qualitative study. Patient Relat Outcome Meas 2012;3:39-49 [FREE Full text] [doi: 10.2147/PROM.S34681] [Medline: 23185121]
13. Haggerty JL. Ordering the chaos for patients with multimorbidity. Br Med J 2012 Sep 7;345:e5915. [doi: 10.1136/bmj.e5915] [Medline: 22960377]
14. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133 [FREE Full text] [Medline: 25642490]
15. Miller WR, Lasiter S, Bartlett Ellis R, Buelow JM. Chronic disease self-management: a hybrid concept analysis. Nurs Outlook 2015 Mar;63(2):154-161 [FREE Full text] [doi: 10.1016/j.outlook.2014.07.005] [Medline: 25241136]
16. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: A meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635 [FREE Full text] [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
17. 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;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
18. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van DG, Wei E. 'You get reminded you're a sick person': personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
19. Bos-Touwen I, Schuurmans M, Monninkhof EM, Korpershoek Y, Spruit-Bentvelzen L, Ertugrul-van der Graaf I, et al. Patient and disease characteristics associated with activation for self-management in patients with diabetes, chronic obstructive pulmonary disease, chronic heart failure and chronic renal disease: a cross-sectional survey study. PLoS One 2015 May 7;10(5):e0126400 [FREE Full text] [doi: 10.1371/journal.pone.0126400] [Medline: 25950517]
20. Siek KA, Ross SE, Khan DU, Haverhals LM, Cali SR, Meyers J. Colorado care tablet: the design of an interoperable personal health application to help older adults with multimorbidity manage their medications. J Biomed Inform 2010 Oct;43(5 Suppl):S22-S26 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.007] [Medline: 20937480]
21. Doyle J, Murphy E, Kuiper J, Smith S, Hannigan C, Jacobs A, et al. Managing Multimorbidity: Identifying Design Requirements for a Digital Self-Management Tool to Support Older Adults with Multiple Chronic Conditions. In: Conference on Human Factors in Computing Systems. 2019 May 4 Presented at: CHI'19; May 7-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300629]
22. Ongwere T, Cantor G, Ramirez MS, Shih P, Clawson J, Connolly K. Design Hotspots for Care of Discordant Chronic Comorbidities: Patients' Perspectives. In: 10th Nordic Conference on Human Computer Interaction. 2018 Sep 29 Presented at: NordiCHI'18; June 23-25, 2018; Oslo, Norway p. 571-583. [doi: 10.1145/3240167.3240221]
23. Lim C, Berry A, Hartzler A, Hirsch T, Carrell D, Bermet Z, et al. Facilitating Self-reflection About Values and Self-care Among Individuals With Chronic Conditions. In: Conference on Human Factors in Computing Systems. 2019 May 01 Presented at: CHI'19; September 7-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300885]
24. Caldeira C, Gui X, Reynolds TL, Bietz M, Chen Y. Managing healthcare conflicts when living with multiple chronic conditions. Int J Hum Comput Stud 2021 Jan;145:102494. [doi: 10.1016/j.ijhcs.2020.102494]
25. Berry A, Lim C, Hartzler A, Hirsch T, Wagner E, Ludman E, et al. How Values Shape Collaboration Between Patients With Multiple Chronic Conditions and Spousal Caregivers. In: Conference on Human Factors in Computing Systems. 2017 May 06 Presented at: CHI'17; June 2-5, 2017; Denver, Colorado,. [doi: 10.1145/3025453.3025923]
26. Lim CY, Berry ABL, Hirsch T, Hartzler AL, Wagner EH, Ludman EJ, et al. Understanding What Is Most Important to Individuals with Multiple Chronic Conditions: A Qualitative Study of Patients' Perspectives. J Gen Intern Med 2017 Aug 28;32(12):1278-1284. [doi: 10.1007/s11606-017-4154-3]

27. Berry AB, Lim CY, Hartzler AL, Hirsch T, Ludman E, Wagner EH, et al. It's good to know you're not a stranger every time. Proc ACM Hum-Comput Interact 2017 Dec 06;1(CSCW):1-20. [doi: 10.1145/3134658]
28. Doyle J, Murphy E, Smith S, Hannigan C, Kuiper J, Jacobs A, et al. Addressing Medication Management Challenges for Older People With Multimorbidities: a Multi-stakeholder Approach. In: EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 May 1 Presented at: Pervasive Health'17; January 6-8, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154883]
29. Tighe SA, Ball K, Kensing F, Kayser L, Rawstorn JC, Maddison R. Toward a digital platform for the self-management of noncommunicable disease: systematic review of platform-like interventions. J Med Internet Res 2020 Oct 28;22(10):e16774. [doi: 10.2196/16774]
30. 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]
31. Czaja S, Boot W, Charness N, Rogers W. Designing for Older Adults: Principles and creative human factors approaches (Third Edition). Florida, United States: CRC Press; Feb 05, 2019.
32. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
33. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 6;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
34. Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: qualitative study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e22488 [FREE Full text] [doi: 10.2196/22488] [Medline: 33464216]
35. Global Status Report on Non-communicable Diseases 2010: Description of the Global Burden of NCDS, Their Risk Factors and Determinants. The World Health Organisation. 2011. URL: https://www.who.int/nmh/publications/ncd_report2010/en/ [accessed 2021-03-01]
36. Doyle J, Murphy E, Hannigan C, Smith S, Bettencourt SJ, Dinsmore J. Designing Digital Goal Support Systems for Multimorbidity Self-management - Insights From Older Adults and Their Care Network. In: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 May 21 Presented at: PervasiveHealth'18; June 7-9, 2018; New York, USA. [doi: 10.1145/3240925.3240982]
37. Doyle J, Hoogerwerf EJ, Kuiper J, Desideri L, Fiordelmondo V, Hannigan C, et al. Whitepaper on the Needs and Requirements of Older People With Multiple Conditions to Self Manage Their Health. ProAct. 2019. URL: http://proact2020.eu/pdf/deliverables/D1.5_Whitepaper-Needs-and-requirements.pdf [accessed 2021-03-01]
38. STARFIELD B, SHI L, MACINKO J. Contribution of primary care to health systems and health. Milbank Q 2005 Sep;83(3):457-502. [doi: 10.1111/j.1468-0009.2005.00409.x]
39. Singh S, Harrison S, Williams JE, Hudson N, Steiner M, Morgan M, et al. How do informal self-care strategies evolve among patients with chronic obstructive pulmonary disease managed in primary care? A qualitative study. COPD 2014 Feb:257-263. [doi: 10.2147/copd.s52691]
40. Russell S, Ogunbayo OJ, Newham JJ, Heslop-Marshall K, Netts P, Hanratty B, et al. Qualitative systematic review of barriers and facilitators to self-management of chronic obstructive pulmonary disease: views of patients and healthcare professionals. NPJ Prim Care Resp Med 2018 Jan 17;28(1):1-15. [doi: 10.1038/s41533-017-0069-z]
41. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-Care Technologies in HCI. ACM Trans Comput-Hum Interact 2015 Dec 14;22(6):1-45. [doi: 10.1145/2803173]
42. Bidwell JT, Vellone E, Lyons KS, D'Agostino F, Riegel B, Juárez-Vela R, et al. Determinants of heart failure self-care maintenance and management in patients and caregivers: a dyadic analysis. Res Nurs Health 2015 Aug 20;38(5):392-402. [doi: 10.1002/nur.21675]
43. Web Content Accessibility Guidelines (WCAG) 2.0. W3C Recommendation. 2008. URL: https://www.w3.org/TR/WCAG20/ [accessed 2021-08-05]
44. Mobile accessibility guidelines. BBC. URL: https://www.bbc.co.uk/accessibility/forproducts/guides/mobile [accessed 2021-08-05]
45. Inclusive Design Research Centre (IDRC). URL: https://achecker.ca/ [accessed 2021-08-05]
46. Contrast checker. WebAIM. URL: https://webaim.org/resources/contrastchecker/ [accessed 2021-08-05]
47. Petrie H, Bevan N. The evaluation of accessibility, usability and user experience. In: Stephanadis C, editor. The Universal Access Handbook. Florida, United States: CRC Press; Jun 13, 2017.
48. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De BI, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
49. The Irish Longitudinal study on Ageing (TILDA). Irish Social Science Data Archive (ISSDA). URL: https://www.ucd.ie/issda/data/tilda/ [accessed 2021-08-05]

50. Deparis S, Tommasi P, Pascale A, Rifai H, Doyle J, Dinsmore J. Building a Risk Model for the Patient-centred Care of Multiple Chronic Diseases. In: IEEE International Conference on Bioinformatics and Biomedicine. 2019 Nov 18 Presented at: BIBM'19; March 23-26, 2019; San Diego, USA. [doi: 10.1109/bibm47256.2019.8983235]
51. Dinsmore J, Hannigan C, Smith S, Murphy M, Kuiper J, O'Byrne E, et al. Digital health platform for integrated and ProACTive patient centred care (ProACT): Protocol for an action-research proof of concept trial. JMIR Preprints. Preprint posted online on July 05, 2020. [doi: 10.2196/preprints.22125]
52. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
53. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London, UK: Taylor and Francis; Jun 11, 2006.
54. Suh H, Shahriaree N, Hekler E, Kientz J. Developing and Validating the User Burden Scale: a Tool for Assessing User Burden in Computing Systems. In: Conference on Human Factors in Computing Systems. 2016 May 07 Presented at: CHI'16; March 7-8, 2016; San Jose, USA. [doi: 10.1145/2858036.2858448]
55. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How Can eHealth Technology Address Challenges Related to Multimorbidity? Perspectives from Patients with Multiple Chronic Conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
56. Mamykina L, Mynatt E, Davidson P, Greenblatt D. MAHI: Investigation of social scaffolding for reflective thinking in diabetes management. In: Conference on Human Factors in Computing Systems. 2008 Apr 05 Presented at: CHI'08; January 6-8, 2008; Florence, Italy p. 477-486. [doi: 10.1145/1357054.1357131]
57. Blondon K, Klasnja P. Designing Supportive Mobile Technology for Stable Diabetes. In: International Conference of Design, User Experience, and Usability. 2013 Jul 21 Presented at: DUEU'13; June 6-9, 2013; Las Vegas, USA p. 361-370. [doi: 10.1007/978-3-642-39241-2_40]
58. Visser L, Shahid S, Mahmud A. Point-of-care Testing for Diabetes Patients: Investigating Diabetes Management by Older Adults. In: Conference on Human Factors in Computing Systems. 2014 Apr 26 Presented at: CHI'14; July 17-19, 2014; Toronto, Canada p. 1645-1650. [doi: 10.1145/2559206.2581193]
59. Tenendez H, Ferrario M, McNaney R, Whittle J. Respiratory Self-care: Identifying Current Challenges and Future Potentials for Digital Technology to Support People With Chronic Respiratory Conditions. In: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 May 20 Presented at: PervasiveHealth'19; 2019; Trento, Italy p. 129-138. [doi: 10.1145/3329189.3329190]
60. Burgess E, Reddy M, Davenport A, Laboi P, Blandford A. Tricky to Get Your Head Around?: Information Work of People Managing Chronic Kidney Disease in the UK. In: Conference on Human Factors in Computing Systems. 2019 May 04 Presented at: CHI;19; May 6-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300895]
61. Gronvall E, Verdezoto N. Understanding Challenges and Opportunities of Preventive Blood Pressure Self-monitoring at Home. In: 31st European Conference on Cognitive Ergonomics. 2013 Aug 26 Presented at: ECCE'13; June 10-11, 2013; Toulouse, France. [doi: 10.1145/2501907.2501962]
62. Pols J. Knowing Patients: Turning Patient Knowledge into Science. Sci Technol Hum Sci 2013 Sep 13;39(1):73-97. [doi: 10.1177/0162243913504306]
63. Craven M, Selvarajah K, Miles R, Schnadelbach H, Massey A, Vedhara K, et al. User Requirements for the Development of Smart-phone Self-reporting Applications in Healthcare. In: International Conference on Human Computer Interaction. 2013 Jul 21 Presented at: HCI'13; September 28-30, 2013; Las Vegas, USA p. 36-45. [doi: 10.1007/978-3-642-39262-7_5]
64. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behaviour change in HCI research. In: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, United States: Association for Computing Machinery; 2011 May Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7 - 12, 2011; BC, Vancouver, Canada p. 3063-3072. [doi: 10.1145/1978942.1979396]
65. Monitoring and Evaluating Digital Health Interventions: a Practical Guide to Conducting Research and Assessment. The World Health Organisation. URL: https://www.who.int/reproductivehealth/publications/mhealth/digital-health-interventions/en/ [accessed 2021-03-01]
66. Nunes F, Fitzpatrick G. Understanding the Mundane Nature of Self-care: Ethnographic Accounts of People Living With Parkinson's. In: Conference on Human Factors in Computing Systems. 2018 Apr 21 Presented at: CHI'18; July 15-17, 2018; Montreal, Canada. [doi: 10.1145/3173574.3173976]
67. D'Haeseleer I, Gerling K, Schreurs D, Vanrumste B, Abeele V. Ageing Is Not a Disease: Pitfalls for Acceptance of Self-management Health Systems Supporting Healthy Ageing. In: 21st International ACM SIGACCESS Conference on Computers and Accessibility. 2019 Oct 28 Presented at: ASSETS'19; May 6-7, 2019; Pittsburgh, USA p. 286-298. [doi: 10.1145/3308561.3353794]
68. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
69. Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]

70. Lee C, Coughlin JF. Perspective: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 3;32(5):747-759. [doi: 10.1111/jpim.12176]
71. Kim S, Gajos K, Muller M, Grosz B, editors. Acceptance of Mobile Technology by Older Adults: a Preliminary Study. In: 18th International Conference on Human Computer Interaction with Mobile Devices and Service. 2016 Sep 06 Presented at: MobileHCI'16; June 2-4, 2016; Florence, Italy. [doi: 10.1145/2935334.2935380]
72. Aamodt IT, Lycholip E, Celutkiene J, von Lueder T, Atar D, Falk RS, et al. Self-care monitoring of heart failure symptoms and lung impedance at home following hospital discharge: longitudinal study. J Med Internet Res 2020 Jan 7;22(1):e15445 [FREE Full text] [doi: 10.2196/15445] [Medline: 31909717]
73. Clawson J, Pater J, Miller A, Mynatt E, Mamykina L. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp'15; July 7-9, 2015; Osaka, Japan p. 647-658. [doi: 10.1145/2750858.2807554]
74. Advancing proactive digital integrated care. ProACT. URL: http://www.proact2020.eu/ [accessed 2021-08-05]

Abbreviations

ARC: action research cycle
CABIE+: context-aware brokering and inference engine
CHD: chronic heart disease
CHF: congestive heart failure
COPD: chronic obstructive pulmonary disorder
EU: European Union
KITE: Knowledge InTEgration
PoC: proof-of-concept
SIMS: subject information management system
SUS: system usability scale
TILDA: The Irish Longitudinal Study on Ageing

Edited by G Eysenbach; submitted 20.07.20; peer-reviewed by N Verdezoto, T Andersen, B Ainsworth, C Caldeira; comments to author 17.11.20; revised version received 31.03.21; accepted 30.05.21; published 15.12.21.

Please cite as:
Doyle J, Murphy E, Gavin S, Pascale A, Deparis S, Tommasi P, Smith S, Hannigan C, Sillevis Smitt M, van Leeuwen C, Lastra J, Galvin M, McAleer P, Tompkins L, Jacobs A, M Marques M, Medina Maestro J, Boyle G, Dinsmore J
A Digital Platform to Support Self-management of Multiple Chronic Conditions (ProACT): Findings in Relation to Engagement During a One-Year Proof-of-Concept Trial
J Med Internet Res 2021;23(12):e22672
URL: https://www.jmir.org/2021/12/e22672
doi:10.2196/22672
PMID:34914612

Original Paper

Design and Development of a Suite of Intimate Partner Violence Screening and Safety Planning Web Apps: User-Centered Approach

Patricia O'Campo[1*], PhD; Alisa Velonis[1,2], MPH, PhD; Pearl Buhariwala[1*], MA, MPH; Janisha Kamalanathan[1*], MSW; Maha Awaiz Hassan[1*], MBBS; Nicholas Metheny[1,3*], PhD, MPH, RN

[1]MAP Center for Urban Health Solutions, St. Michael's Hospital, Toronto, ON, Canada

[2]Division of Community Health Sciences, University of Illinois at Chicago School of Public Health, Chicago, IL, United States

[3]School of Nursing and Health Studies, University of Miami, Coral Gables, FL, United States

[*]these authors contributed equally

Corresponding Author:
Patricia O'Campo, PhD
MAP Center for Urban Health Solutions
St. Michael's Hospital
30 Bond St
Toronto, ON, M5W 1W8
Canada
Phone: 1 4766304000
Email: patricia.ocampo@unityhealth.to

Abstract

Background: The popularity of mobile health (mHealth) technology has resulted in the development of numerous apps for almost every condition and disease management. mHealth and eHealth solutions for increasing awareness about, and safety around, intimate partner violence are no exception. These apps allow women to control access to these resources and provide unlimited, and with the right design features, safe access when these resources are needed. Few apps, however, have been designed in close collaboration with intended users to ensure relevance and effectiveness.

Objective: The objective of this paper is to discuss the design of a suite of evidence-based mHealth and eHealth apps to facilitate early identification of unsafe relationship behaviors and tailored safety planning to reduce harm from violence including the methods by which we collaborated with and sought input from a population of intended users.

Methods: A user-centered approach with aspects of human-centered design was followed to design a suite of 3 app-based safety planning interventions.

Results: This review of the design suite of app-based interventions revealed challenges faced and lessons learned that may inform future efforts to design evidence-based mHealth and eHealth interventions.

Conclusions: Following a user-centered approach can be helpful in designing mHealth and eHealth interventions for marginalized and vulnerable populations, and led to novel insights that improved the design of our interventions.

(J Med Internet Res 2021;23(12):e24114) doi:10.2196/24114

KEYWORDS

intimate partner violence; web-based applications; women; user-centered design

Introduction

In recent years, web-based and mobile health (mHealth) technology has expanded greatly, providing a unique space for individuals to access information and resources to manage and improve health and well-being [1-3]. With upward of 5 billion individuals subscribing to wireless mobile services globally [4], digital space allows those seeking information or support to do so in a user-controlled, discreet, and accessible way [5-7]. This technology fills a unique gap for survivors of intimate partner violence (IPV), giving control to individuals themselves to access information about safety and resources. mHealth tools can be used at any time and any place, and survivors can return to the information as many times as needed to explore their

concerns and their options as their situations change. In the past 20 years, more than 300 mobile or web-based apps for IPV have emerged [3,8], but only few have achieved widespread adoption. One limitation of apps that have failed to gain traction is their inability to meet the needs of specific communities or targeted groups, including the identification of local and appropriate resources.

The Partner Violence Implementation Science group is a collaboration between researchers, service providers, and survivors of IPV; our work is focused on cocreating interventions supporting individuals' safety and self-determination. As part of these efforts, we identified the need for electronic IPV screening and safety planning resources tailored to the communities in which we work [3,4]. While a number of apps are available to assist women with identifying IPV and formulating safety plans, our community partners were concerned that many did not adequately address the nonlinear and multiphase process many women go through in an effort to increase safety for themselves and their children [9-11]. Gaining awareness of the possible dangers involved with taking action is one of these challenges, and women who have experienced IPV often lament that this awareness—along with information about how to increase relationship safety—did not come earlier in the process [12-14]. While our initial intent was to develop a tool to aid health care providers screen for IPV in an outpatient orthopedic clinic, our community partners—including survivors of IPV—felt that a web-based mHealth tool that was easy and quick to use, exclusively focused on identifying "red flags" for abusive behaviors, and that could help users identify safety options tailored for their situation was a critical need across settings. While we recognize that IPV is experienced by persons of all genders, our research focused on the experiences of the largest group of those experiencing IPV: cisgender women who are in relationships with cisgender men.

First, we looked at a variety of screening apps aimed at helping women identify if they are in an abusive relationship. After previewing a variety of publicly available screening apps across North America, we discovered that these tools often left out 1 or more types of violence, or did not use validated screening questions to detect IPV [15,16]. This led to our decision to create our own screening app that could help a wide range of individuals recognize patterns of unsafe behaviors in their own or others' relationships and immediately access information about abuse and local services.

Next, for the smaller subset of women who may be experiencing behaviors related to IPV, a separate app was needed that could help them assess the severity of their situation (including the risk for lethality) and would promote safety planning behaviors tailored for their situations. One existing app provided much of what we were looking for. Developed by researchers at Johns Hopkins University [17-19] and tailored for Canadian audiences by researchers at Western University [20,21], the MyPlan app is an evidence-based mHealth app aimed at increasing users' understanding of potentially dangerous patterns in their relationships reducing their ambivalence, or decisional conflict, about acting to alter these patterns and has been found to reduce future IPV in international randomized control trials [17,20,22-24]. After engaging with our target population and consultation with the teams at Johns Hopkins University and Western University, we decided to build on this evidence base and strong foundation to address additional needs present in our populations of interest as identified by our community partners in the Greater Toronto Area (GTA): women who were looking for more information about IPV, the majority of whom are not currently experiencing any unsafe relationship behaviors, and a second population of women who are living with IPV and need information about the potential for serious consequences of the abuse (up to and including lethality) and access to tailored safety planning that references geographically and culturally relevant strategies and resources.

This paper describes the rigorous process we engaged in to develop 3 evidence-based apps, each of which was developed for a particular audience and tailored to their specific needs. Using elements of human-centered design (HCD) [25], we ensure the end users (women at risk for IPV and those experiencing IPV) are involved in all stages of design and testing of our suite of mHealth tools. This includes iterative elicitation regarding the quality of experience, feelings of safety using the apps, and perceived helpfulness of the tools themselves throughout the design and testing processes [26,27]. Very few IPV screening or safety planning apps provide detailed information about the development or testing of the product, and there is scant peer-reviewed evidence regarding the process of developing the app, user testing, and evaluation [8]. Because of the critical role these tools can play for women who are concerned about the safety of themselves or a loved one, ensuring that the end product both addresses the primary concerns of women in a local community and uses evidence-informed processes to provide knowledge or bring about change is essential. Thus, we begin to shed light on this process by describing our iterative, community-engaged approach to building 3 separate, yet complementary, IPV mHealth interventions: an IPV screening tool (WithWomen), an individualized safety planning web app (Pathways), and the rapid adaptation of this app to the realities of living with IPV during COVID-19 (Promoting Safety in Emergencies, or PROMiSE). This suite of apps is summarized in Table 1.

Methods

Approach

Our approach to the design and adaptation of our suite of apps relied on the formation of partnerships with health care providers working to implement a screening program into their outpatient clinic setting, as well as service providers at IPV shelters, counseling groups for women with lived experience of IPV, and a peer support network for women living with IPV. We felt it important to keep these perspectives present throughout our design and adaptation process. Our partners played many roles: they advised us on key design features and content, connected us with women with lived experience who would provide feedback on the screening questions and functionality of the web apps, informed us of resources within the region which should be listed in the web apps, and connected us with plain language experts to ensure the content was appropriate for our population. The process used to develop each of these is

discussed below. All 3 apps were approved by the St. Michael's Hospital Research Ethics Board (REB # 15-361). Table 1 outlines the components of our suite of apps, their target populations, and main innovations, while Table 2 reviews the demographic characteristics of samples used to develop and test each of the mHealth tools.

The demographic characteristics of those who participated in the various surveys and interviews to inform the development of our screening and safety decision-support app are presented in Table 2.

Table 1. Partner Violence Implementation Science app suite components.

App feature	WithWomen Screener	Pathways	PROMiSE[a]
Target population	Women in relationships with men	Women in male–female relationships who have moderate-to-high safety concerns in their relationships	Women in male–female relationships who have moderate-to-high safety concerns in their relationships during public health emergencies
Purpose	Screen for potential IPV[b]	Safety planning and local resource connection	Modified safety planning and local resource connection during public health emergencies
Year of release	2018	2019	2020

[a]PROMiSE: Promoting Safety in Emergencies.

[b]IPV: intimate partner violence.

Table 2. Demographic characteristics of research participants[a].

Demographics	Screening app			Pathways app		PROMiSE[b] app
	Cognitive interviews (n=18), n (%)	Anonymous encounter surveys (n=16), n (%)	App user testing (n=41), n (%)	Staff: preliminary user testing (n=19), n (%)	Clients: user testing (n=46), n (%)	User testing (n=7), n (%)
Age, years						
16-34	5 (28)	9 (56)	18 (44)	8 (42)	11 (24)	3 (43)
35-55+	13 (72)	7 (44)	22 (54)	11 (58)	35 (76)	4 (57)
Born in Canada and Indigenous						
Yes: Indigenous	0 (0)	1 (6)	4 (10)	4 (20)	8 (17)	3 (43)
Yes	10 (56)	9 (50)	17 (41)	8 (42)	16 (35)	0 (0)
No	8 (44)	7 (44)	20 (49)	10 (53)	21 (46)	4 (57)
Experiences of IPV[c] in the past 5 years by the participant or someone close to them						
Yes	9 (50)	Not asked	31 (76)	13 (68)	39 (85)	2 (28)
No	8 (44)	Not asked	10 (24)	6 (32)	7 (15)	5 (72)

[a]Numbers do not always total to 100% due to missing responses for selected categories.

[b]PROMiSE: Promoting Safety in Emergencies.

[c]IPV: intimate partner violence.

Developing the WithWomen Screening App

Selecting the Screening Questions

We sought to have valid and reliable screening questions. After reviewing multiple validated IPV screening tools and the peer-reviewed literature [28], we selected the *Hurt, Insult, Threaten, and Scream* (HITS) instrument to serve as the base for our screening app. The HITS instrument is brief, has been used in general practice and emergency department settings [29,30], demonstrated acceptable sensitivity (88%) and specificity (range 86%-97%), and has been extensively tested within different populations (eg, tested in women/men, Hispanic and African American women as well as in Spanish) [31,32]. However, one of the limitations of HITS is its inability to identify subtle experiences of IPV, such as coercive and controlling behaviors, as well as experiences of sexual IPV. We identified several additional questions by consulting the peer-review literature on validated screening questions concerning coercive control and through conversations with our community partners (eg, "partner controlling what you wear").

To be sure we were identifying those app users with the greatest need for immediate assistance, we limited the recall period to the past 12 months and asked questions in terms of frequency, rather than only identifying if behaviors had happened at all. Additionally, we intentionally asked about both short-term (eg, dating) and long-term (eg, partners) relationships.

Testing the Questions

To ensure our screening questions were acceptable and appropriate, we sought input from a cross-section of women.

We conducted anonymous encounter surveys with women in local shopping malls. We approached women in the food courts, explained our research, and asked if they had 10 minutes to complete a short survey. We asked women to give us feedback on the clarity and acceptability of the questions and whether we were missing questions about a particular type of violence. We also consulted with women who were more likely to have lived experience of IPV by posing the same questions to female clients who accessed services at our partner community agencies (eg, women's shelters, organizations serving women involved with the justice system). We also consulted service providers who serve women more generally such as at health clinics or agencies who help women with employment or housing to seek their input.

Next, as we narrowed the set of candidate questions, we conducted 3 cycles of cognitive interviews to ensure the clarity of the questions [33]. These were face-to-face surveys asking women to tell us in their own words what each question was asking and to get their advice on any terms or phrases that were hard to understand or inappropriate. We revised the questions based upon feedback we received in between each cycle. We also sought to test the reliability of our questions by administering the questions to a subset of women and contacting them 1 month later to retake the survey. Agreement between test and retest was assessed through computing the intraclass correlation coefficient (ICC). We also established evidence for convergent validity with questions capturing safety-related activities such as talking to a counselor about relationship concerns or missing work due to relationship issues [34].

The final screening instrument consisted of 9 items. The next step was to assign a level of risk to an individual's summed score on the full screening instrument. To do this, we first assigned a value of risk for the answers to each question. Using information on the severity of the violence described in the question and corroborating it with existing severity ratings for similar questions [17,32], we assigned risk values to each question, with the lowest value of risk being 0 and the highest level of risk being 3. For example, if the respondent's partner, ex-partner, or someone the respondent dated insulted her frequently, she received a question-specific level of risk of 2. On the physical violence item, she automatically received a question-specific score of 3 if she experienced any amount of beating, punching, kicking, strangling, or harm with a weapon. A sum score of 0-4 across all 9 questions indicates few to no safety concerns, 5-8 indicates there may be safety concerns, and 9-24 indicates that there are moderate to high safety concerns in the relationship (see Textbox 1 for scoring summary).

Finalizing Our Screening Questions

Using feedback from the cognitive interviews, we made slight wording changes to our questions to ensure their acceptability to women. For example, the question on sexual violence originally featured the word "coerce"; however, not all women interviewed understood the meaning behind the word and it was changed to "pressure, threaten, or force." Overall, the questions were highly acceptable and clear to potential users, largely due to the fact that we relied on previously validated screening tools as the source of our questions. Almost perfect test–retest agreement was demonstrated via the 16 reliability interviews (ICC 96%; 95% CI 90%-99%). Reliability was similar when the sample was reduced to only those women who reported at least one positive answer to the screening questions (ICC 95%; CI 79%-99%).

Textbox 1. WithWomen final rapid intimate partner violence screening questions with points assigned to answers to reflect level of safety risk in a relationship.

1. Over the last 12 months, how often did you feel uncomfortable doing or saying things around your current partner or someone you're currently dating?

0=never, rarely, sometimes

1=frequently

2. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated INSULT you or talk down to you?

0=Never, rarely

1=Sometimes

2=Frequently

3. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated yell, shout, or curse at you?

0=Never, rarely

2=Sometimes

3=Frequently

4. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated control who you see, where you go, what you do, or what you wear?

0=Never

2=Rarely, Sometimes

3=Frequently

5. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated make you feel afraid or scared of them?

0=Never

2=Rarely, sometimes

3=Frequently

6. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated THREATEN to harm you or someone you care about?

0=Never

2=Rarely

3=Sometimes, Frequently

7. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated physically HURT you?

0=Never

2=Rarely

3=Frequently, Sometimes

8. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated beat, punch, kick, strangle, or hurt you with a weapon?

0=Never

3=Rarely, Sometimes, Frequently

9. Over the last 12 months, how often did your partner, an ex-partner, or someone you dated force, threaten, or pressure you to participate in any sexual activity when you didn't want to?

0=Never

3=Rarely, Sometimes, Frequently

Points for each question are summed and categorized into the following categories. The results screen displays the color (eg, yellow) and the explanation of the result (ie, "there are some things about your relationship that are of concern") at the end of rapid screening.

0-4: Healthy (Green): There are few to no concerns regarding safety in your relationship

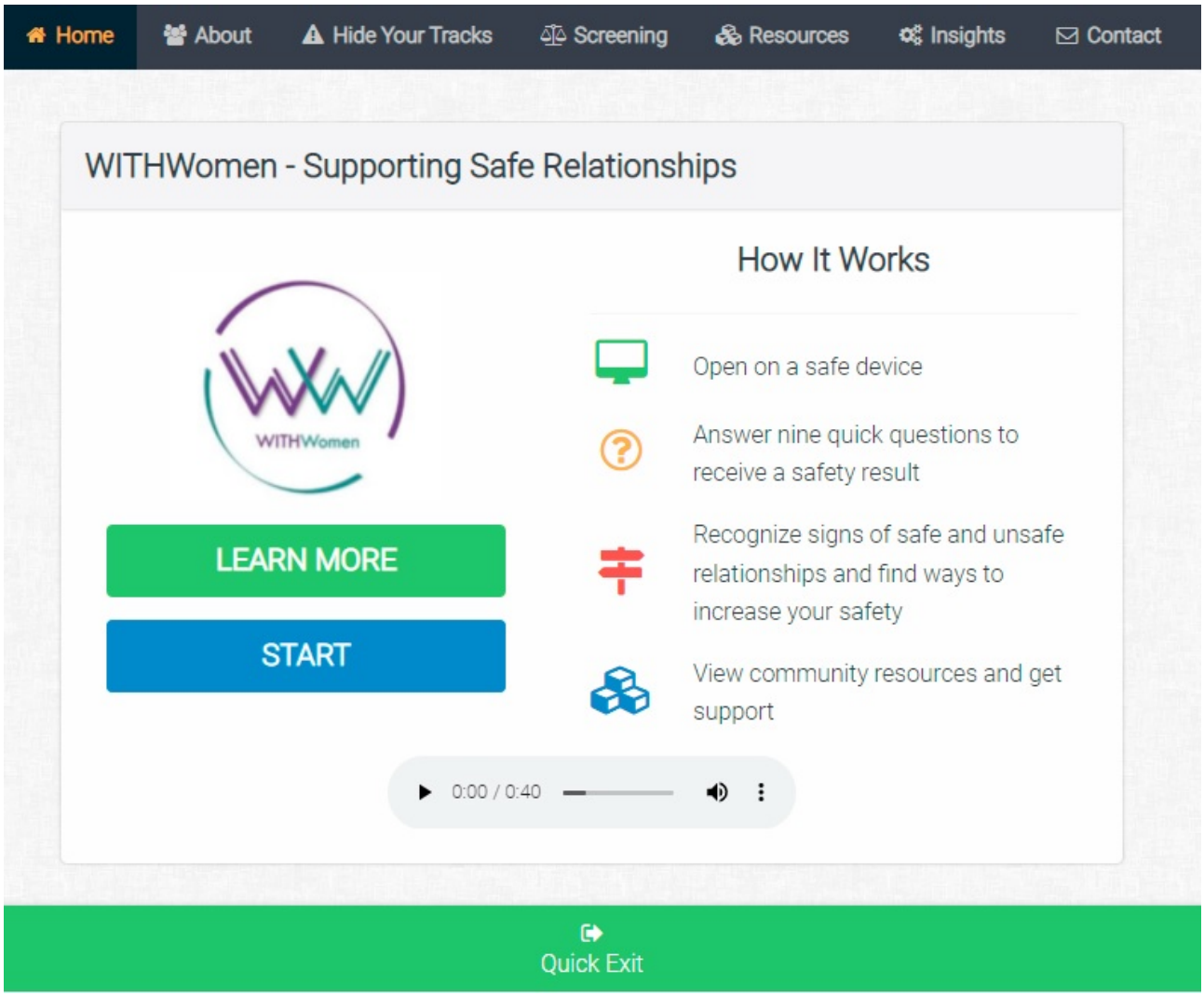
5-8: Caution (Yellow): There are some things about your relationship that are of concern.

9-24: Confirmed abuse (Red): Your relationship has many safety concerns.

We established concurrent validity of our scales by correlating scores on the 9-item violence scale with responses to questions about whether women took part in activities that might be expected for women who were concerned about safety in their relationships: searched the internet for information about IPV; talked to a social worker or other professional or family and

friends about her relationship; missed work because of relationship issues; or called the police due to relationship issues. Spearman correlations between the scores on the IPV screener and the 4 behaviors ranged from 0.72 to 0.82. We did ask about 1 more behavior, sought medical attention due to violence, and its correlation with our IPV screener was lower (0.57). Thus, concurrent validity was moderately high for all but 1 of the items, seeking medical care, but given that this activity happens relatively rarely and is only expected for 1 type of violence, physical violence, this level of correlation might be expected. The WithWomen landing page is shown in Figure 1.

Figure 1. WithWomen Landing Page.



The user journey for our WithWomen Screening app is presented in Figure 2, illustrating the simplicity and completeness of the app as it screens for IPV, provides an interpretation of the screening result, and provides links to resources related to prevention or management of IPV. Moreover, the app can be used to serve the public or be used in a clinic setting providing opportunities to share the screening result with clinic staff.

Figure 2. WithWomen User Journey.

Developing the Pathways Safety Planning App

Overview

As mentioned earlier, existing safety planning and decision-support apps did not meet the unique needs of our target population. In particular, we learned from our work with our community that the current apps required changes in terms of language, levels of interactivity, more flexibility in the user journey, accessibility of populations with literacy challenges, content, and safety features, and thus a new app was required. Since Pathways is intended to be a tool specifically for users who have self-identified as currently or recently experiencing IPV (ideally because they have used the WithWomen app), we designed Pathways to introduce users to the several options

available in the app to begin making safety planning decisions. These options include (1) assessing severity, danger, and potential lethality of violent behaviors in their relationship; (2) identifying their safety priorities; (3) learning about safety actions to take that align with their priorities; and (4) learning about healthy and unhealthy relationships. We drew on existing safety planning and decision-support apps for IPV and made key modifications for our target population of urban Canadian women. These included tailoring the app to address women's top concerns related to safety planning while still in or recently exiting a violent relationship. The goal of this app, then, is to allow women to prioritize their safety planning needs and receive tailored support, tangible action items, and referrals to local resources that match their highest concerns.

Step 1: Prototype Development

To develop the prototype for our app, we recruited 19 service providers and 31 women who identified as IPV survivors to interact with sections of the existing MyPlan app. In addition to this, we asked a smaller subset of participants with lived experience to review the full app online. Participants were recruited from 8 IPV-focused organizations in greater Toronto. We completed this user research over the course of 2 months, making note of their preferences, experiences with interacting with the existing app content and activities, and suggested modifications to inform the development of our Pathways app.

We then drew on 2 sources of data to determine the specific concerns most important to our users, which likely differed from the college-age females for whom the original MyPlan app was intended. We recruited 16 additional women who did not participate in the assessment of the existing safety planning app with different levels of lived experiences of IPV. Participants completed an anonymous survey that asked them to select how important 10 concerns are to women who are experiencing IPV as they make important decisions about their safety. In addition to those considered in the existing app, we reviewed the literature and consulted service provider key informants to arrive at these 10 concerns, which included having resources (finances, housing, legal services), privacy, immigration, and career-related concerns [13,35] (Table 3). When literacy was a challenge, a trained data collector supported respondents by reading out the instructions, explaining the activity, and writing down their responses. These answers were summed across participants and priorities that received the highest ratings were chosen as priority areas to which women could tailor their safety planning activities in Pathways.

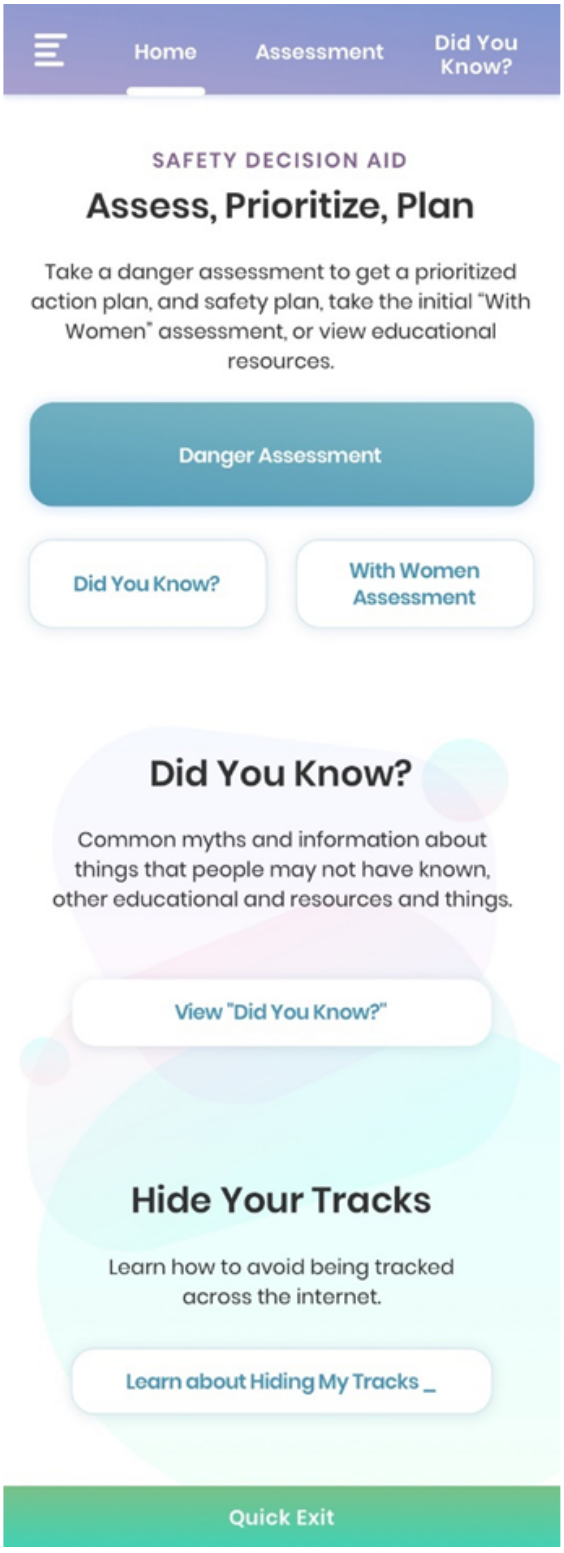
Table 3. Relative ranking for each priority area for the My Concerns section of Pathways.

Rank	Safety priorities	Relative ranking
1	Having resources (finances, housing, legal support)	High
2	The health and well-being of someone close to you	High
3	Housing concerns	High
4	Privacy	High
5	Your personal health and well-being	Moderate
6	Language barriers	Moderate
7	Studies and career	Moderate
8	Immigration concerns	Moderate
9	Feelings for a partner	Low
10	Connections to the community	Low

Step 2: User Testing

Once the prototype data were collected, we worked with a software development firm, Tactica Interactive [36], to develop a working prototype of Pathways. Using this prototype and following HCD principles, we measured user experience with the app, looking specifically at functionality, ease of navigation, and the comprehensiveness of each of the tailored concern sections [26]. This process involved asking IPV survivors and service providers to engage with an online version of the app and provide feedback on usability of the different sections in an anonymous survey. A total of 26 individuals, including 5 IPV service providers, spent approximately 30-40 minutes reviewing all components of the app. User research was conducted at 4 IPV service organizations. One of the organizations provided trained interpreters to non-English speaking clients to make group activities more inclusive. As interpreters were already available in this setting, 5 women who spoke languages other than English were able to participate by working with an interpreter and a trained data collector to complete their survey responses. The user experience with Pathways is outlined in Figure 3.

Figure 3. Pathways Landing Page.



Step 3: App Refinement

Finalizing Content for the Pathways App

For the section on user's priorities, data from surveys conducted with participants and open-ended questions on what might be missing from app content shared with those taking part in the user research provided many options for what we might consider including in the prioritization section of the app.

Linking the safety options to a user's priorities has been demonstrated to reduce decisional conflict about IPV safety planning [17,19,22]. In an effort to more closely align users' concerns with their safety plans, "Resources" was further divided into "Finances," "Housing," and "Legal Support" as separate categories of Resource-based concerns. We removed "Child's well-being" from the priority setting activity and instead incorporated a safety plan focused on children as a

separate section in the Pathways app, where it could be given more prominence for those who select it.

The final 5 categories of concerns selected for inclusion in Pathways were each linked to an action plan, which suggests 3-6 specific actions users can take to address each area of concern. Safety steps associated with lower-ranked priorities were added into the optional viewing sections of the app. For example, suggestions to improve "Privacy" were included in the stalking-related safety plan, and also in the "Hide Your Tracks" section, which includes instructions for safe web browsing, while suggestions to address immigration concerns were added into the "Legal Support" section of the app. Algorithms were added to Pathways such that safety plans for the user's most highly rated priority were presented to the user first.

Finalizing Pathways

Overall, user testing interviews confirmed the usefulness, understandability, appropriateness, and comprehensiveness of Pathways for our population. No major changes to safety planning–related content were suggested by respondents or our advisory groups. However, some common themes emerged about the language and structure of the app. For example, suggestions were made to simplify language, reduce text, and use check boxes whenever possible. In some sites, words such as "cue," "priority," and "prioritize" were not considered common language. We simplified the language and changed section titles (eg, "My Priorities" became "My Concerns" in Pathways). Several clients and service providers commented about the need to use reaffirming, rather than alarming language. Service providers also acknowledged the need to balance between providing validation and affirmations without normalizing severely dangerous behaviors. This was particularly relevant to the Danger Assessment section of the app. Our team removed questions that were in the Danger Assessment section but not used in the ratings of violence severity and prefaced the questions with plain language content explaining the rationale for asking them.

Several participants discussed the benefit of incorporating more tailored resources based on various needs of diverse populations. The recommendations were mostly around community supports specifically for minorities and people with language barriers, culturally sensitive services, and legal advice. Keeping in mind that language, transit access, and cost are barriers to accessing formal supports, we included links to support services that could be accessed by people free of charge or on sliding scale without the need for referrals or health insurance. In many cases, we listed support services in Pathways that are offered in multilingual, multicultural, or multisite service settings, and we included options to search resources by postal code whenever possible to allow women to find the most appropriate resources.

Another key area of concern pertained to the privacy, security, and accessibility of the app. To address these concerns, we incorporated a quick exit bar at the bottom of each app page that allows users to quickly hide their screen by redirecting them to the Google search engine. We also added a separate section in the app called "Online safety" that contains detailed information on secure web browsing. Our inclusive design intern redesigned the "Did You Know" section of our app, which originally described aspects of healthy and unhealthy relationships, into an activity where women interactively explore violence-related information to increase their knowledge about relationship violence. To increase accessibility for users with visual impairments or low literacy, we added an audio feature to each page of the app. The feature reads all text aloud and narrates navigation options on each screen.

The majority of the Pathways beta testers provided positive feedback (good or very good experience using the app) and noted that they would recommend the app to someone else. The lowest scored section of our survey was in relation to ease of navigation, with 15/46 (33%) of respondents identifying sections of the app that had too much text. To address these concerns, we took these sections to plain writing workshops hosted by our hospital's patient education department, where plain language experts reviewed the material and provided feedback. We then hired an editor to incorporate suggestions and reduced the amount of text in flagged sections of the app by 20%. We also simplified the user flow, by moving certain sections of text to separate pages, and making those pages optional to view.

Data for Our Apps

For safety reasons and also to honor concerns around privacy expressed by our informants with lived experience throughout the process of creating this suite of apps, we designed our apps to collect all data anonymously. Thus, no names, IP addresses, or any other identifiers are collected or stored. The research team does, however, collect other data for quality improvement purposes (eg, distribution of scores for screening or the danger assessment or priority concerns of those using Pathways or what time of day the apps are accessed), but we examine all data anonymously. We also ensured that no trace is left on any devices that have accessed any of our apps. The data are stored on servers in North America and the research team owns the data.

Developing PROMiSE

The Pathways app was launched in December 2019. However, by early 2020 it was evident that the COVID-19 pandemic and the policies required to mitigate its spread (eg, stay-at-home orders, reduced service capacity) would lead to increased time spent in close quarters with abusive partners, widespread uncertainty, and financial strain, all stressors known to increase the risk for IPV [37]. Rates of IPV have indeed increased since the advent of the pandemic [38,39], and many women are unable to access IPV services due to reduced capacity and increased demand, making IPV screening and safety planning mHealth interventions critical to maximizing women's safety during this and future public health emergencies. However, we received feedback from community partners that aspects of the Pathways app were not well suited to public health emergencies. For instance, the decision-support tools in Pathways are centered on seeking help outside the home and connecting with community services, many of which are either not open due to government restrictions, have closed due to financial strain, or are over capacity due to increased demand. It became clear that a safety planning tool was needed that responded to the realities women face during public health emergencies and that provides

up-to-date, relevant information, and advises on actions to maximize safety in a safe, discreet way.

Using the same HCD-informed approach, we began a rapid research project to adapt Pathways for use during COVID-19 that involved (1) conducting a rapid systematic search of the peer-reviewed and gray literature on strategies that women experiencing IPV in the context of COVID-19 might find helpful and (2) convening an expert panel of IPV survivors and IPV service providers in the GTA to brainstorm new and modified strategies that women experiencing IPV in the context of COVID-19 might find helpful. This 3-month rapid research yielded 22 strategies that were either highly or somewhat recommended (eg, staying connected with others and planning for safety) and 6 that were not recommended (eg, hiding items that might be used as weapons) as they might make violence worse. Armed with new information about how women who are currently experiencing violence can maximize their safety during COVID-19, we connected again with Tactica to develop the new PROMiSE app from the existing Pathways app.

In addition to updating the content, we felt it was especially important to ensure the look and feel of the app was discreet, given that many more women may be using PROMiSE in close proximity to their abusers. To this end, we partnered with Tactica to develop a disguise feature for PROMiSE, wherein the app content is overlaid onto an innocuous webpage. We chose the main Pinterest board for Home and Garden Television as our innocuous landing page based on feedback from survivors of IPV that this would be unlikely to arouse suspicion from an abuser. When a woman visits the site, she will recognize the PROMiSE logo in the top right corner from advertisements and marketing materials (Figure 4). When this is clicked, the PROMiSE content will appear as a “pop-up” (Figure 5). In addition to the quick exit bar present in WithWomen and Pathways, users who click anywhere outside the “pop-up” box that contains the PROMiSE content will hide the app itself, revealing only the background website. The first time a woman visits the site, she will be offered an interactive tutorial that explains how to hide the content and make it reappear using hotkeys (desktop version) or a hotspot (on touchscreen devices).

Figure 4. PROMiSE Landing Page as Disguised.

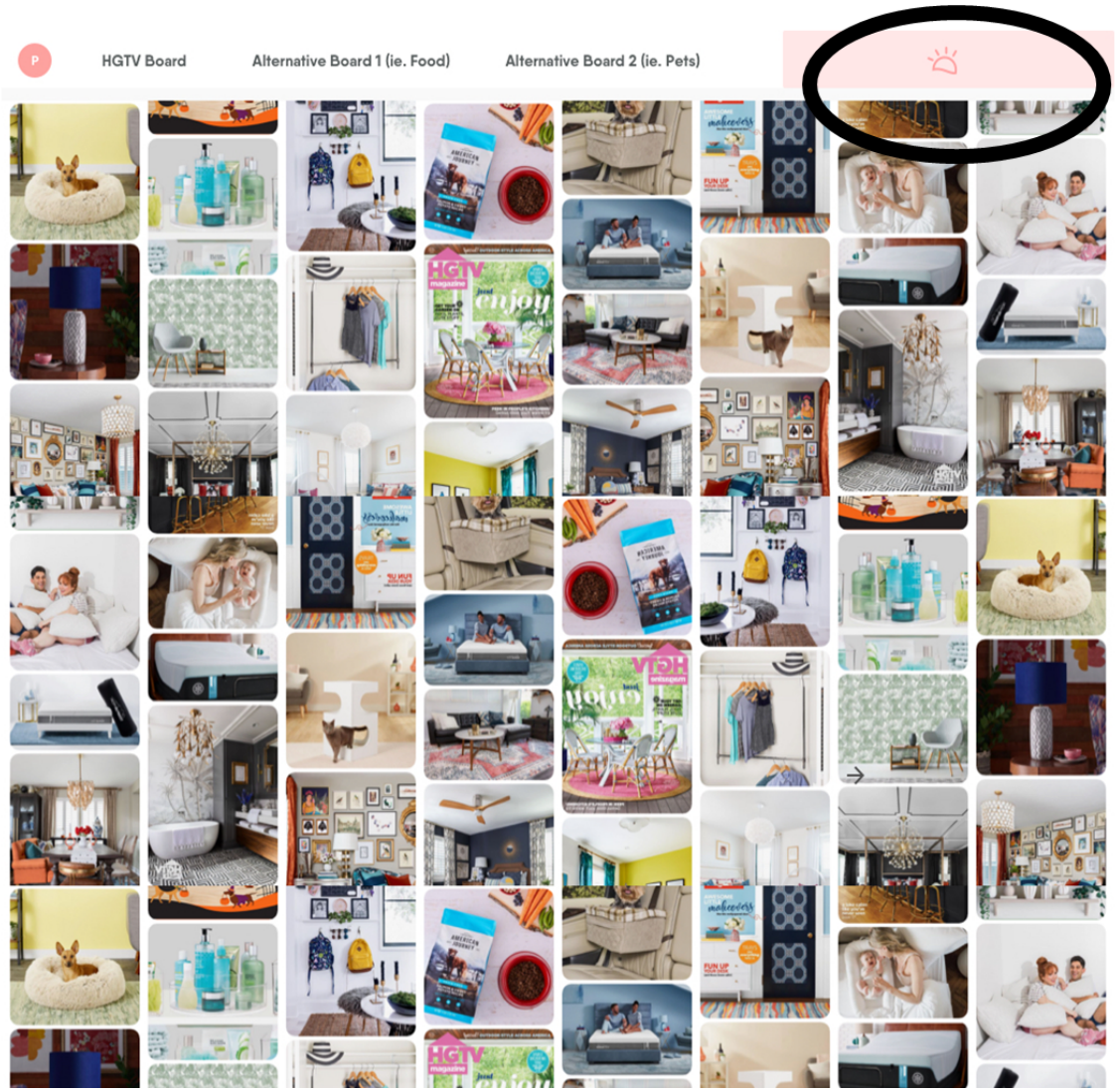
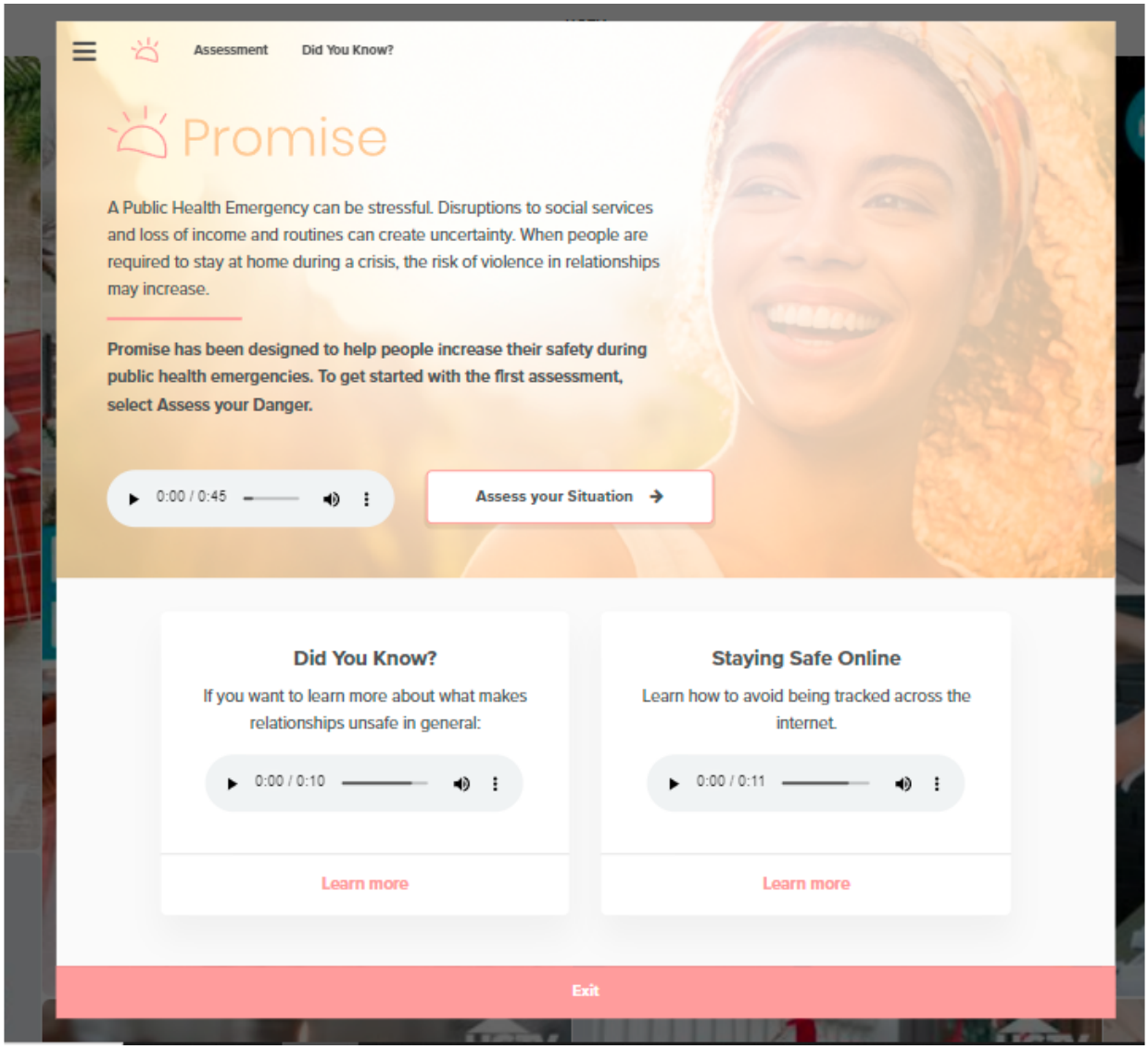


Figure 5. The actual PROMiSE landing page.



Because the layout and technical aspects of PROMiSE are identical to those of Pathways, we forewent much of the technical user testing in service of a rapid rollout of this app in the midst of the pandemic. However, once a prototype of PROMiSE was available, we invited women with lived experience of IPV and IPV service providers to navigate the app and provide feedback on its new look and feel and to identify any glitches that may have occurred in the creation of the app (n=7; Table 2). This information will be incorporated into later updates of PROMiSE. PROMiSE was launched on December 1, 2020

Discussion

Lessons Learned

Web-based and mHealth apps are increasingly common for IPV [1,3]. This paper demonstrates how we built on the existing set of IPV screening and safety planning web-based apps to tailor and create a suite of tools appropriate for our local population of women who are at risk of or needing to plan for safety around IPV before and during COVID-19. Furthermore, given the range of preventive and educational activities taken up by web-based and mHealth IPV-related apps, we focused our efforts on 2 particular challenges: (1) helping women to learn early on about unsafe relationship behaviors; and (2) once violence is confirmed, tailored safety planning to maximize safety. These apps were designed to complement the vast set of existing resources and services in our metropolitan area but also serve to provide a unique resource by taking into account a woman's particular priorities and being accessible any time of day and any place that the internet is available.

Systematic reviews of mHealth, computer-, and web-based apps have noted the scarcity of engagement with affected communities in the design of applications that impact the relevance and utilization of such apps [3]. We addressed this limitation in our work by using elements of HCD, which sought input from and cocreation with survivors of IPV and health care providers who work closely with this population. While this required that additional time be built into each of our phases of research, this increased the relevance of our apps to our target population. We found this to be critically important given the stage of IPV that we are targeting with these apps: for WithWomen screening we are targeting women who are yet unaware of the safety threats present in their relationships, and

for Pathways and PROMiSE we are targeting women with varying levels of violence in their intimate relationships who are needing to take some kind of action to increase safety in their relationships. While an existing app [18] was the template for our Pathways and PROMiSE, our apps' new features requested and endorsed by intended users included priorities that aligned with their preferences, interactive sections on healthy relationship information, freedom within the user journey to access any part of the app, plain language text, and an audio option.

We learned from survivors, providers, and also from the literature that abuse itself impacts the survivors' trust in others and confidence in themselves for making sound decisions about their relationships and safety-related actions. Furthermore, despite the vast strength and resilience of survivors, their experiences with relationships where they are routinely admonished and denigrated by perpetrators result in pervasive survivor self-doubt and self-blame, further affecting information processing and decision making around safety [40-44]. Our user research and advice from our partners helped shape the language and tone of the text in the apps. Because of the engagement of women and health and social services providers in the process of undertaking user research, we employed affirmative and strength-based language and phrases. Moreover, given levels of self-doubt among survivors who are living with abuse, we also used language that simultaneously prioritized trusting users' own instincts about maintaining safe behaviors while also encouraging safety planning action. One issue we were not able to overcome is the impact of missing data in our screening scales. If a respondent skips 1 or more questions, a value of 0 is used for the skipped question, potentially artificially deflating actual level risk.

Working closely with partners had other benefits. Our fracture clinic partner adopted a technology-enhanced approach to screening for IPV using our WithWomen IPV screening app. We used implementation science strategies to assist them with developing a suitable screening protocol that minimally impacted their busy clinic schedule [45] and for the first time ever the clinic had access to data about screening and IPV prevalence. While we did not plan for the broader use of our apps in service or health care settings, virtual visits with clients have significantly increased due to the pandemic and along with it the opportunity to incorporate technology-enhanced screening or safety planning during or in between client visits. With the release of PROMiSE, with its focus on safety planning during public health emergencies and enhanced safety features, we have been asked by providers to assist with the creation of options for including our apps within their virtual interactions with clients.

Because the key features of MyPlan that led to positive safety planning outcomes were retained in our Pathways app, we feel that another clinical trial demonstrating that use of the app over usual care is not warranted [19]. However, we are planning to use innovative single-case experimental study designs to demonstrate, on a much smaller sample of participants, that the PROMiSE app is safe to use when in the home and generates positive impacts on IPV knowledge and safety planning activities [46].

While we had evidence-based models to draw from, we discovered that creating these apps took far longer than we anticipated for several reasons. The iterative nature and multiple cycles of developing prototypes for sections of the apps, gaining feedback on those, and making modification to the prototypes take time. Our research team had to build capacity in some areas such as learning about the privacy concerns when our initial designs sought to link our rapid IPV screening app to the hospital data system and understanding the technical aspects of discreet app design. As we relied on the expert consultants and our hospital IT department in a set of iterative conversations about the app features that are or are not aligned with privacy requirements (eg, finding a secure way to link our app that is housed on external servers with the hospital-based electronic medical records), our process was further slowed as we were not a priority of the hospital's IT concerns. Finally, we also experienced significant delays due to the initial technology development partner we had chosen. Our team initially engaged the services of a laboratory connecting clinicians with biomedical and computer engineering students to apply technological solutions to the real-world problems faced by clinicians in their practice. We were eager to give a student team the opportunity to build our app while simultaneously cutting down on development costs to accommodate our shoestring budget. Unfortunately, the combination of our lack of experience with developing an app, our inability to communicate the technological specifications to the student team, and our lack of understanding of how short or long the development of these apps should take resulted in a slow process that spanned almost 18 months that ultimately ended in failure. We probably erred by not including a member from the technology team in our discussions of the research process; however, rapid turnover in the student team would have posed another challenge to that strategy. Fortunately, when we subsequently turned to a professional app developer the work was completed in a matter of weeks. While we were fortunate enough to engage a Master's student from the Inclusive Design program at the Ontario College of Arts and Design as an intern to advise on graphic design and user experience for WithWomen and Pathways, a diminishing overall budget precluded allowing us to have a designer or more advanced features to these aspects of our suite. Separate funding for PROMiSE, however, allowed us to make use of a professional designer and more technical consultants, resulting in increased functionality and advanced privacy features.

Conclusion

With the use of technology in health and social service care intervention and delivery becoming more commonplace, there is an opportunity to develop high-quality IPV screening apps to implement in practice. The aim of this venture was to create a suite of IPV screening and safety planning apps for use with local women that contain relevant information and resources in a safe, discreet way. Evidence and user testing feedback have indicated what women want and need from IPV screening and safety planning apps: that is, the product to be relevant to its user population with easy navigation, HCD features, and the ability to access discreetly.

Acknowledgments

This research was supported by a generous donation to the St. Michael's Hospital Foundation. We thank Drs Nancy Glass and Marilyn Ford-Gilboe for their expert input and advise on the creation of our safety decision-support web-based app. We also thank James Case for sharing information about the technological details of MyPlan and for advice on creating our app. Drs Janice du Mont and Robin Mason of Women's College Hospital, Toronto, provided valuable insight and feedback on the development of PROMiSE. Finally, we thank the women who generously spent time providing feedback on early versions of our apps, in particular, the women from Peer2Peer, a survivor-led organization seeking to help women living with or a history of IPV in greater Toronto.

Conflicts of Interest

None declared.

References

1. Goldstein KM, Zullig LL, Dedert EA, Alishahi Tabriz A, Brearly TW, Raitz G, et al. Telehealth Interventions Designed for Women: an Evidence Map. J Gen Intern Med 2018 Dec;33(12):2191-2200 [FREE Full text] [doi: 10.1007/s11606-018-4655-8] [Medline: 30284173]
2. Haddad SM, Souza RT, Cecatti JG. Mobile technology in health (mHealth) and antenatal care-Searching for apps and available solutions: A systematic review. Int J Med Inform 2019 Jul;127:1-8. [doi: 10.1016/j.ijmedinf.2019.04.008] [Medline: 31128820]
3. Anderson EJ, McClelland J, Meyer Krause C, Krause KC, Garcia DO, Koss MP. Web-based and mHealth interventions for intimate partner violence prevention: a systematic review protocol. BMJ Open 2019 Aug 10;9(8):e029880. [doi: 10.1136/bmjopen-2019-029880] [Medline: 31401604]
4. GSM Association. 2017 Jun 13. URL: https://www.gsma.com/newsroom/press-release/number-mobile-subscribers-worldwide-hits-5-billion/ [accessed 2021-01-15]
5. Aryana B, Brewster L. Design for mobile mental health: Exploring the informed participation approach. Health Informatics J 2020 Jun;26(2):1208-1224 [FREE Full text] [doi: 10.1177/1460458219873540] [Medline: 31566470]
6. Covolo L, Ceretti E, Moneda M, Castaldi S, Gelatti U. Does evidence support the use of mobile phone apps as a driver for promoting healthy lifestyles from a public health perspective? A systematic review of Randomized Control Trials. Patient Educ Couns 2017 Dec;100(12):2231-2243. [doi: 10.1016/j.pec.2017.07.032] [Medline: 28855063]
7. Haggerty LA, Hawkins JW, Fontenot H, Lewis-O'Connor A. Tools for screening for interpersonal violence: state of the science. Violence Vict 2011;26(6):725-737. [doi: 10.1891/0886-6708.26.6.725] [Medline: 22288092]
8. Eisenhut K, Sauerborn E, García-Moreno C, Wild V. Mobile applications addressing violence against women: a systematic review. BMJ Glob Health 2020;5(4):e001954 [FREE Full text] [doi: 10.1136/bmjgh-2019-001954] [Medline: 32399255]
9. Burke JG, Mahoney P, Gielen A, McDonnell KA, O'Campo P. Defining appropriate stages of change for intimate partner violence survivors. Violence Vict 2009;24(1):36-51. [doi: 10.1891/0886-6708.24.1.36] [Medline: 19297884]
10. Catallo C, Jack SM, Ciliska D, Macmillan HL. Identifying the turning point: using the transtheoretical model of change to map intimate partner violence disclosure in emergency department settings. ISRN Nurs 2012;2012:239468 [FREE Full text] [doi: 10.5402/2012/239468] [Medline: 22792480]
11. Jane, Stoever. Transforming Domestic Violence Representation. Kentucky Law Journal 2013 Oct 31;101(3):483-537 [FREE Full text] [doi: 10.1093/acprof:osobl/9780199989591.003.0017]
12. Daoud N, Matheson FI, Pedersen C, Hamilton-Wright S, Minh A, Zhang J, et al. Pathways and trajectories linking housing instability and poor health among low-income women experiencing intimate partner violence (IPV): Toward a conceptual framework. Women Health 2016;56(2):208-225. [doi: 10.1080/03630242.2015.1086465] [Medline: 26358378]
13. Daoud N, Matheson FI, Pedersen C, Hamilton-Wright S, Minh A, Zhang J, et al. Pathways and trajectories linking housing instability and poor health among low-income women experiencing intimate partner violence (IPV): Toward a conceptual framework. Women Health 2016;56(2):208-225. [doi: 10.1080/03630242.2015.1086465] [Medline: 26358378]
14. DeKesseredy, W, Dragiewicz M, Schwartz M. Abusive Endings: Separation and Divorce Violence Against Women. Oakland, CA: University of California Press; 2017.
15. Purple Evolution Incorporated. 2017. URL: https://www.facebook.com/PevoApp/ [accessed 2019-08-21]
16. Soujourner Peace App. 2019. URL: https://familypeacecenter.org/resources/online-resources/ [accessed 2019-08-21]
17. 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]
18. Johns Hopkins University School of Nursing. MyPlan App. 2021. URL: https://www.myplanapp.org/ [accessed 2021-01-15]

19. Glass N, Eden KB, Bloom T, Perrin N. Computerized aid improves safety decision process for survivors of intimate partner violence. J Interpers Violence 2010 Nov;25(11):1947-1964 [FREE Full text] [doi: 10.1177/0886260509354508] [Medline: 20040709]
20. 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]
21. Ford-Gilboe M, Varcoe C, Scott-Storey K, Perrin N, Wuest J, Wathen CN, et al. Longitudinal impacts of an online safety and health intervention for women experiencing intimate partner violence: randomized controlled trial. BMC Public Health 2020 Feb 26;20(1):260 [FREE Full text] [doi: 10.1186/s12889-020-8152-8] [Medline: 32098633]
22. 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]
23. Hegarty K, Tarzia L, Murray E, Valpied J, Humphreys C, Taft A, et al. Protocol for a randomised controlled trial of a web-based healthy relationship tool and safety decision aid for women experiencing domestic violence (I-DECIDE). BMC Public Health 2015 Aug 01;15:736 [FREE Full text] [doi: 10.1186/s12889-015-2072-z] [Medline: 26231225]
24. 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]
25. International Standards Organization. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. ISO Standards Catalogue. 2010. URL: https://www.iso.org/standard/52075.html [accessed 2021-12-07]
26. Martínez-Pérez B, de LTI, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. J Med Syst 2013 Oct;37(5):9976. [doi: 10.1007/s10916-013-9976-x] [Medline: 24014255]
27. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-Centered Design Study: Enhancing the Usability of a Mobile Phone App in an Integrated Falls Risk Detection System for Use by Older Adult Users. JMIR Mhealth Uhealth 2017 May 30;5(5):e71 [FREE Full text] [doi: 10.2196/mhealth.7046] [Medline: 28559227]
28. Rabin RF, Jennings JM, Campbell JC, Bair-Merritt MH. Intimate partner violence screening tools: a systematic review. Am J Prev Med 2009 May;36(5):439-445.e4 [FREE Full text] [doi: 10.1016/j.amepre.2009.01.024] [Medline: 19362697]
29. Spangaro J, Vajda J, Klineberg E, Lin S, Griffiths C, Saberi E, et al. Intimate partner violence screening and response in New South Wales emergency departments: A multi-site feasibility study. Emerg Med Australas 2020 Aug;32(4):548-555. [doi: 10.1111/1742-6723.13452] [Medline: 31965691]
30. Chen P, Rovi S, Washington J, Jacobs A, Vega M, Pan K, et al. Randomized comparison of 3 methods to screen for domestic violence in family practice. Ann Fam Med 2007;5(5):430-435 [FREE Full text] [doi: 10.1370/afm.716] [Medline: 17893385]
31. Chan C, Chan Y, Au A, Cheung G. Reliability and Validity of the “Extended - Hurt, Insult, Threaten, Scream” (E-Hits) Screening Tool in Detecting Intimate Partner Violence in Hospital Emergency Departments in Hong Kong. Hong Kong Journal of Emergency Medicine 2017 Dec 11;17(2):109-117. [doi: 10.1177/102490791001700202]
32. Sherin KM, Sinacore JM, Li XQ, Zitter RE, Shakil A. HITS: a short domestic violence screening tool for use in a family practice setting. Fam Med 1998;30(7):508-512. [Medline: 9669164]
33. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-238. [doi: 10.1023/a:1023254226592] [Medline: 12769135]
34. Wolford-Clevenger C, Kuhlman S, Elledge LC, Smith PN, Stuart GL. A Preliminary Validation of the Suicidal Behavior Exposure Scale. Psychol Violence 2019 Jul;9(4):442-450 [FREE Full text] [doi: 10.1037/vio0000170] [Medline: 31934496]
35. Buel S. Fifty obstacles to leaving, aka Why Women Stay. The Colorado Lawyer 1999;28(10):19-28.
36. Tactica International. URL: https://tactica.ca/?doing_wp_cron=1629130143.1195750236511230468750 [accessed 2021-01-15]
37. Jewkes R. Intimate partner violence: causes and prevention. Lancet 2002 Apr 20;359(9315):1423-1429. [doi: 10.1016/S0140-6736(02)08357-5] [Medline: 11978358]
38. Godin M. 2020 As cities around the world go on lockdown, victims of domestic violence look for a way out. Time Magazine. New York, NY: Time USA, Inc; 2020. URL: https://time.com/5803887/coronavirus-domestic-violence-victims/ [accessed 2021-01-14]
39. Dubinski K, Margison A. National survey finds domestic violence during pandemic was more frequent and severe. Canadian Broadcasting Corporation. 2020 Aug 25. URL: https://www.cbc.ca/news/canada/london/national-survey-domestic-violence-end-violence-now-anova-1.5698028 [accessed 2021-01-15]
40. Matheson FI, Daoud N, Hamilton-Wright S, Borenstein H, Pedersen C, O'Campo P. Where Did She Go? The Transformation of Self-Esteem, Self-Identity, and Mental Well-Being among Women Who Have Experienced Intimate Partner Violence. Womens Health Issues 2015;25(5):561-569. [doi: 10.1016/j.whi.2015.04.006] [Medline: 26116987]
41. Brosi M, Rolling E, Gaffney C, Kitch B. Beyond Resilience: Glimpses into Women’s Posttraumatic Growth after Experiencing Intimate Partner Violence. The American Journal of Family Therapy 2019 Nov 25;48(1):1-15 [FREE Full text] [doi: 10.1080/01926187.2019.1691084]

42. Anderson KM, Renner LM, Danis FS. Recovery: resilience and growth in the aftermath of domestic violence. Violence Against Women 2012 Nov;18(11):1279-1299. [doi: 10.1177/1077801212470543] [Medline: 23334815]
43. Lim BHP, Valdez CE, Lilly MM. Making Meaning Out of Interpersonal Victimization: The Narratives of IPV Survivors. Violence Against Women 2015 Sep;21(9):1065-1086. [doi: 10.1177/1077801215590670] [Medline: 26091685]
44. Ulloa E, Hammett JF, Guzman M, Hokoda A. Psychological growth in relation to intimate partner violence: A review. Aggression and Violent Behavior 2015 Nov;25:88-94 [FREE Full text] [doi: 10.1016/j.avb.2015.07.007]
45. Velonis AJ, O'Campo P, Rodrigues JJ, Buhariwala P. Using implementation science to build intimate partner violence screening and referral capacity in a fracture clinic. J Eval Clin Pract 2019 Jun;25(3):381-389. [doi: 10.1111/jep.13128] [Medline: 30932286]
46. O'Campo P, Mutaner C, Metheny N, Yakubovich A. Improving experimental designs for social science research: refinements to the single case experimental design. Grant funded by the Social Sciences and Humanities Research Council (SSHRC). Social Sciences and Humanities Research Council of Canada. 2019. URL: https://maphealth.ca/sced-evaluation/ [accessed 2021-01-09]

Abbreviations

GTA: Greater Toronto Area
HCD: human-centered design
HITS: Hurt, Insult, Threaten, and Scream
ICC: intraclass correlation coefficient
IPV: intimate partner violence
mHealth: mobile health
PROMiSE: Promoting Safety in Emergencies

Edited by R Kukafka; submitted 04.09.20; peer-reviewed by K McDonnell, N Glass; comments to author 20.10.20; revised version received 12.12.20; accepted 10.08.21; published 21.12.21.

Please cite as:
O'Campo P, Velonis A, Buhariwala P, Kamalanathan J, Hassan MA, Metheny N
Design and Development of a Suite of Intimate Partner Violence Screening and Safety Planning Web Apps: User-Centered Approach
J Med Internet Res 2021;23(12):e24114
URL: https://www.jmir.org/2021/12/e24114
doi:10.2196/24114
PMID:34931998

Original Paper

Effectiveness and Moderators of an Internet-Based Mobile-Supported Stress Management Intervention as a Universal Prevention Approach: Randomized Controlled Trial

David Daniel Ebert[1], BSc, MSc, PhD; Marvin Franke[2], BSc, MSc; Anna-Carlotta Zarski[2], BSc, MSc, PhD; Matthias Berking[2], BSc, MSc, PhD; Heleen Riper[3], BSc, MSc, PhD; Pim Cuijpers[3], BSc, MSc, PhD; Burkhardt Funk[4], BSc, MSc, PhD; Dirk Lehr[4], BSc, MSc, PhD

[1]Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany

[2]Clinical Psychology and Psychotherapy, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

[3]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[4]Department of Health Psychology and Applied Biological Psychology, Leuphana University of Lueneburg, Lueneburg, Germany

Corresponding Author:
David Daniel Ebert, BSc, MSc, PhD
Department of Sport and Health Sciences
Technical University of Munich
Georg-Brauchle-Ring 60/62
TUM Campus D
Munich, 80992
Germany
Phone: 49 15116781057
Email: david.daniel.ebert@tum.de

Abstract

Background: Emerging evidence indicates the effectiveness of internet-based mobile-supported stress management interventions (iSMIs) in highly stressed employees. It is yet unclear, however, whether iSMIs are also effective without a preselection process in a universal prevention approach, which more closely resembles routine occupational health care. Moreover, evidence for whom iSMIs might be suitable and for whom not is scarce.

Objective: The aim of this study was to evaluate the iSMI GET.ON Stress in a universal prevention approach without baseline inclusion criteria and to examine the moderators of the intervention effects.

Methods: A total of 396 employees were randomly assigned to the intervention group or the 6-month waiting list control group. The iSMI consisted of 7 sessions and 1 booster session and offered no therapeutic guidance. Self-report data were assessed at baseline, 7 weeks, and at 6 months following randomization. The primary outcome was perceived stress. Several a priori defined moderators were explored as potential effect modifiers.

Results: Participants in the intervention group reported significantly lower perceived stress at posttreatment (d=0.71, 95% CI 0.51-0.91) and at 6-month follow-up (d=0.61, 95% CI 0.41-0.81) compared to those in the waiting list control group. Significant differences with medium-to-large effect sizes were found for all mental health and most work-related outcomes. Resilience (at 7 weeks, P=.04; at 6 months, P=.01), agreeableness (at 7 weeks, P=.01), psychological strain (at 6 months, P=.04), and self-regulation (at 6 months, P=.04) moderated the intervention effects.

Conclusions: This study indicates that iSMIs can be effective in a broad range of employees with no need for preselection to achieve substantial effects. The subgroups that might not profit had extreme values on the respective measures and represented only a very small proportion of the investigated sample, thereby indicating the broad applicability of GET.ON Stress.

Trial Registration: German Clinical Trials Register DRKS00005699; https://www.drks.de/DRKS00005699

(J Med Internet Res 2021;23(12):e22107) doi:10.2196/22107

KEYWORDS

stress management intervention; universal prevention; occupational health; moderators

Introduction

Occupational stress is a major public health problem, with high prevalence in western countries [1-3]. It is associated with a range of adverse consequences, including an increased risk for coronary disease [4,5], mental health problems [6,7], and sleeping problems [8]. Moreover, stress-related productivity loss, absenteeism, presentism, and health care usage lead to substantial economic costs [9]. Although manifold meta-analytic evidence exists for the effectiveness of psychological face-to-face stress management interventions for employees [10,11], the high prevalence of stress demands for highly cost-effective and scalable solutions. Internet-based mobile-supported stress management interventions (iSMIs) can be a great solution as they can offer several benefits such as (1) their accessibility at any time and place, (2) the possibility for participants to work and review materials at their own pace, and (3) their high potential for scalability [12]. In theory, only a small increase in resources is required for reaching a greater proportion of the eligible population. However, the real costs of iSMIs are greatly linked to the amount of guidance delivered by professional support [13], thereby limiting the possible reach of these interventions. Therefore, unguided interventions are an important puzzle piece to combat the high prevalence of stress.

The most recent meta-analysis on iSMIs found overall significant effectiveness (d=0.43, 95% CI 0.31-0.54) with small-to-medium effect sizes, but substantial heterogeneity between studies [14]. Subgroup analyses found that guided interventions (d=0.64, 95% CI 0.50-0.79; n=7) were significantly more effective than unguided interventions (d=0.33, 95% CI 0.20-0.46; n=18). Yet, the most effective iSMI GET.ON Stress yielded high effect sizes even when delivered in an unguided format [13], with an effect size (d=0.96, 95% CI 0.70-1.21) significantly exceeding that reported in the overall sample. In regard to the often cited replication crisis in Psychology [15] and other fields [16-18], it is crucial to evaluate if these effect sizes can withhold further inspection. Moreover, replicating these studies can help to address their shortcomings and add to the existing literature.

The first shortcoming relates to the evaluation in selected samples. Evidence for the best evaluated iSMI GET.ON Stress [13,19-25] is based on trials in which participants have been preselected based on the high baseline score in perceived stress (Perceived Stress Scale [PSS-10] ≥22) [26,27]. Using a broad range of eligibility criteria, however, can limit the real-life applicability of the results [28], as in routine preventive occupational health care, iSMIs are usually offered to a wide range of participants. Therefore, it is crucial to investigate treatment effects without preselection in a universal prevention approach [29]. This approach has not only theoretical but also practical advantages: while reaching a wider proportion of the working population, it reaches those that benefit from selected and indicated prevention without the cost for screening them specifically [30,31]. Additionally, it allows the inclusion of adults with lower symptom severity who are still motivated to improve their conditions, which might be a better indicator for their readiness for health behavior change [32].

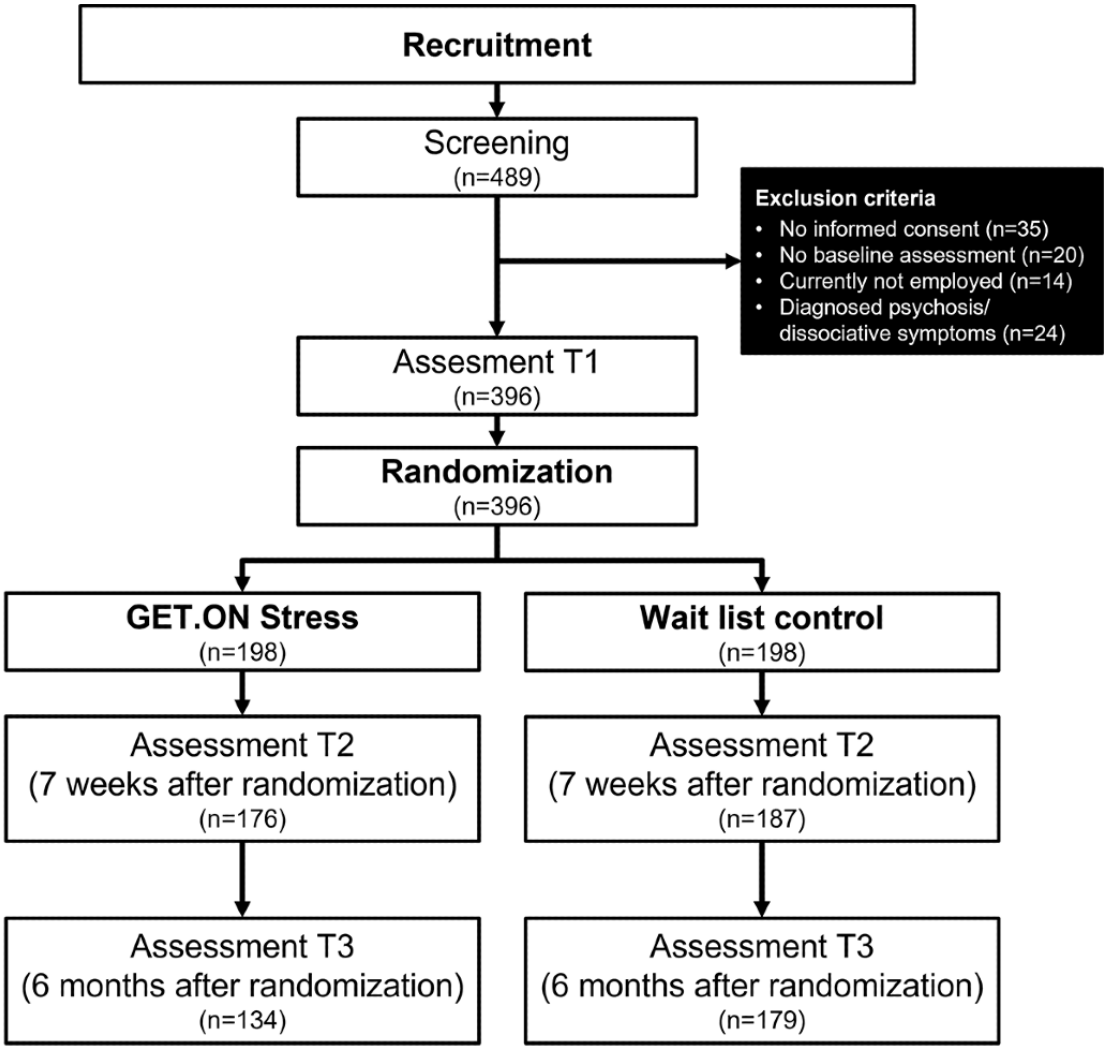
The second shortcoming relates to the lack of moderator analyses. Even though internet-based intervention research is rapidly growing, empirical data on the moderators of internet-based mobile-supported interventions [33-39] and, in particular, internet occupational health interventions are scarce [34]. Furthermore, randomized controlled trials (RCTs) are usually powered to detect effects on the primary outcome and hence, underpowered to reliably test moderator hypotheses [40]. With reported numbers needed to treat of 5.43 for unguided iSMIs (and 4.20 in the overall sample) [14], it is clear, however, that a substantial proportion of the participants do not profit yet from taking part in an intervention. Therefore, identifying the moderators of iSMI effects is crucial for at least 3 reasons: (1) knowing who likely profits from the intervention helps to identify relevant populations, (2) knowing which patients are unlikely to profit from the treatment helps to prevent wasting resources and provides valuable information on where custom-tailoring such interventions to subgroups of patients is necessary, and (3) a better knowledge regarding who is (un)likely to profit from these interventions helps to identify mechanisms of change that are relevant for these interventions [41]. Therefore, the aim of this study was to evaluate the effects of the iSMI GET.ON Stress as a universal prevention approach, without preselecting participants based on symptom severity. Moreover, in an exploratory approach, a broad range of a priori selected potential effect modifiers were tested in a sample adequately powered for moderator analyses.

Methods

Design

A two-armed RCT (N=396) was conducted comparing a self-guided iSMI (GET.ON Stress) condition (intervention group [IG]) and a waiting list control (WLC) condition, with both conditions having full access to treatment as usual. Participants were recruited via the occupational health program of a large health insurance company in Germany (BARMER) in a way that mimics the intended implementation of the intervention in routine practice in the future, that is, by offering the intervention in a public occupational mental health approach to the general working population in a web-based setting. Recruitment was directed at the general working population and not restricted to members of the health care insurance company. Recruitment occurred mainly through reports in the membership magazine of the insurance company, and the insurance company's occupational health management workers informed human resource departments of collaborating small- and medium-sized companies about the possibility for their employees to participate in the trial. Assessments took place at baseline (T1), at posttreatment (7 weeks, T2), and at 6 months (T3; see Figure 1 for a detailed overview of assessments). This study was approved by the University of Marburg ethics committee.

Figure 1. Flow of the study. T1: baseline; T2: at 7 weeks posttreatment; T3: at 6 months posttreatment.



Interested individuals received an email with detailed information about the study procedures and an invitation to take part in the study. If they were interested, they were asked to complete a web-based screening questionnaire. Individuals meeting the eligibility criteria completed the baseline assessment and were invited to complete the informed consent form. Once the full written informed consent was received, participants entered the study and were randomly allocated to 1 of the 2 study conditions. Randomization took place at a ratio of 1:1. An independent third party who did not have any information about the participants performed the allocation. Randomization was carried out using an automated computer-based random integer generator (randlist). Other researchers could not bias the randomization process since participants were randomized in the order of the incoming informed consent form. Participants were not blinded to study conditions. During the randomization process, the allocation was concealed from the participants, researchers involved in recruitment, and the study administration team.

Eligibility Criteria

Inclusion and exclusion criteria were kept to a minimum to mimic routine conditions as closely as possible. Participants were included who (1) were 18 years and older, (2) were currently employed, (3) had internet access, and (4) provided a valid email address. Individuals were only excluded if they were at risk of suicide (indicated by a score of 2 or higher on the Beck Depression Inventory Suicide Item [42]) or if they self-reported to have been diagnosed with psychosis or dissociative symptoms. Besides that, there were no restraints for participants, for example, concerning low or high stress, depression, or other symptom severity criteria.

Intervention

As this study evaluates an intervention that was evaluated in an indicated prevention context before, intervention content and delivery methods mirror those in related studies [20]. The training GET.ON Stress is based on Lazarus' transactional model of stress [43]. According to this model, there are 2 coping strategies to deal with stress. One is a problem-oriented approach in changing the situation. The other approach focusses on situations where only the evaluation of the situation can be adjusted; this is called emotional regulation. Therefore, the training focuses on strategies for systematic problem-solving as well as emotion regulation. The intervention consists of 7 regular sessions and 1 booster session (provided 4 weeks after the last regular session was completed). The program starts with psychoeducation about stress (session 1), which guides participants to understand and find coping strategies for common problem situations. Based on their personal stressors,

participants are asked to record their goals and motivation for the training. This is followed by 2 sessions focusing on problem-solving skills (sessions 2-3) where participants are guided through a 6-step method based on problem-solving therapy [44]. The personal plan will be implemented between the sessions before participants are asked to redo the exercise with the same or a different problem. The next part of the training (sessions 4-6) is based on the Affect Regulation training [45,46]. Each session focusses on 1 emotion regulation strategy: muscle and breathing relaxation, acceptance of emotions, as well as effective self-support with 15-minute audio files guiding users through the different exercises. Participants are encouraged to redo the exercises between sessions on a daily basis. The regular training is completed by a session guiding the creation of a "Plan for the future" (session 7), in which users are asked to review their progress and develop strategies according to their personal stress indicators in the future. Moreover, they are guided to write a letter to their future self. The booster session reviews the essential content of the intervention and allows participants to reassess their goals and plans for the future. Additionally, participants could choose from an array of elective modules in sessions 2 to 6 addressing common stress-related topics such as time management, sleep hygiene, or social support. Participants were advised to complete 1-2 sessions per week.

Across sessions, the training was composed of interactive education, audio and video files, and additional downloadable material. Additionally, testimonials offered impulses and examples concerning the exercises. The intervention is based on responsive design, meaning that it could be completed on a desktop computer, tablet, or smartphone. Through providing multiple choices among various response options, the content of the intervention was automatically tailored to the specific needs and interests of the individual participants. To integrate the newly acquired knowledge into daily life, homework assignments and behavioral planning were a crucial part of the training. Web-based diaries were also offered. The training was not accompanied by therapeutic guidance. However, the participants could opt for automatic text messages on their mobile phones choosing either a light (1 text message every other day) or intensive support (2-3 text messages per day). Text messages focused on ultrabrief exercises to be carried out in daily life routine, aiming to facilitate transfer from training into real life.

Outcomes

All measures were assessed through web-based self-report questionnaires and are described below. The assessment took place on a secured web-based system with 256-bit AES encryption. All outcomes were published a priori in detail in the trial register. The primary outcome was the perceived level of stress as measured by the German version of the PSS-10 [26,27] since it is based on Lazarus' transactional model of stress. The 10 items were answered on a 5-point Likert scale (ranging from 0=never to 4=very often) referring to the past week. Owing to its sum score, the total scale scores range from 0 to 40. Cronbach alpha was .83 at T1, .91 at T2, and .91 at T3 in this study.

Secondary outcomes included measurement of mental health and work-related outcomes. Among those concerning mental health, the following outcomes were measured using specified scales: depression, using the German adaptation of the Center for Epidemiological Studies' Depression Scale (15 items, range 0-45; α=.94) [47,48]; mental health–related quality of life, using the Assessment of Quality of Life (8 items, range 8-41; α=.84) [49]; resilience, using the Connor-Davidson Resilience Scale (10 items, range 10-50; α=.93) [50]; emotional exhaustion, using the subscale of the Maslach-Burnout-Inventory (5 items, range 1-6; α=.93) [51]; mental well-being, using the Warwick-Edinburgh Mental Wellbeing Scale (14 items, range 14-84; α=.96) [52]; and psychological well-being, using the 5-item World Health Organization Well-Being-Index (5 items, range 0-100; α=.91) [53]. The following work-related outcomes were assessed: effort-reward imbalance, using the Effort-Reward-Imbalance Questionnaire-Short Form (10 items, α=.84) [54]; occupational self-efficacy, using the Occupational Self-Efficacy Scale (6 items, range 6-36; α=.93) [55]; work engagement, using the Utrecht Work Engagement Scale (9 items, range 0-6; α=.97) [56]; productivity loss, using the Work-Limitations-Questionnaire (8 items, α=.87) [57]; and absenteeism and presentism days, using the respective questions from the German version of the Trimbos and Institute of Medical Technology Assessment Cost Questionnaire for Psychiatry (2 items) [58].

In addition, the following variables were assessed to be tested as potential moderators: personality, using the Big-Five-Inventory (10 items, subscale range α=.20-.77) [59]; motivation for treatment, using the Psychotherapy Motivation Questionnaire-Short Form (4 items, α=.36) [60]; self-regulation competencies, using the Self-Regulation scale (10 items, range 10-40; α=.84) [61]; and general self-efficacy, using the General Self-Efficacy Scale (10 items, range 10-40; α=.90) [62].

Power Analysis

The sample size of the study was optimized to detect potential moderator effects. Previous studies on the same intervention found effects on the primary outcome of perceived stress at T2 ranging from d=0.84 to d=0.93 [13,23,63] in highly stressed employees. Assuming potentially lower effects, owing to the inclusion of less impaired employees, of approximately d=0.60, we would have needed to include 90 participants. Given that simulation studies showed that the sample size required to detect an interaction effect of the same magnitude needs to be at least threefold to fourfold, compared to the sample sizes needed for the primary analyses [64], we included 396 participants. This sample allowed us not only to detect a small regression effect of f^2 with a power (1-ß) of 80% and an error probability (α) of .05 but also a small intergroup effect size of d=0.25 in a one-tailed test (calculated with G*Power).

Statistical Methods

All analyses are reported following the Consolidated Standards of Reporting Trials guidelines for reporting parallel group randomized trials [65]. Therefore, following the intention-to-treat principle, missing data were handled with multiple imputations [66], using 100 estimations per missing value. Outcome levels of IG and WLC were compared at T2

and T3 by using analysis of covariance (ANCOVA), with baseline levels as covariates. Additionally, Cohen *d* with 95% CIs were calculated based on the imputed data set by comparing means and SDs of the 2 groups at the respective time points. Furthermore, the difference in means along with its 95% CIs was calculated. To improve interpretability, reliable change was calculated according to the method of Jacobson and Truax [67] as an indicator of the number of participants with treatment response. Participants were accredited a treatment response if their PSS-10 score differed more than SD 5.16 in T1-T2 and T1-T3. Besides, the number needed to treat to achieve 1 additional treatment response compared to the control group was calculated. To evaluate the effect of the intervention on the risk for symptom worsening, we also calculated the number of participants with a reliable deterioration of symptoms by using the Reliable Change Index and calculated the absolute and relative reduction of risk, including the respective 95% CIs.

Moderator Analysis

Regression analyses were used to test baseline moderator X intervention condition interaction effects by using the SPSS macro PROCESS [68]. In case of a significant interaction effect, we applied the Johnson-Neyman technique [68,69], which tests the conditional effect of X on Y at different values along the continuum of the moderator and calculates transition points, where the effect changes between statistically significant and not significant (at the α level of significance). This allows to identify a "region of significance" of the effect, giving away at which values of the moderator a significant effect can be found. This warrants essential advantages against the pick-a-point approach in which mostly only 3 arbitrarily chosen values of the moderator are tested since even the widespread use of the mean as well as SD is highly sample-specific and can thus lead to wrong conclusions [68]. Moderation variables were neither standardized nor mean-centered since it makes no difference for the moderation effect [68] and can even harm the interpretation if dichotomous variables are altered before the analysis.

Results

Participants

The study flow is illustrated in Figure 1. All 396 participants provided data at T1. In the IG, 176 (88.8%) participants at T2 and 134 (67.7%) participants at T3 and in the WLC group, 187 (94.4%) participants at T2 and 179 (90.4%) participants at T3 completed the assessment. The demographic characteristics of participants can be found in Table 1. The average age of the participants was 41.76 (SD 10.09) years. The sample was predominantly female (302/396, 76.3%), married, or in a relationship (209/396, 52.8%), as well as highly educated (285/396, 72%). Most participants were employed full-time (296/396, 74.7%), in a permanent employment relationship (306/396, 77.3%), and nearly half of them held a management function (169/396, 42.7%). Only a small portion (35/396, 8.8%) was self-employed. The average working experience was 17.58 (SD 10.36) years and the participants' jobs were in various working sectors, with the majority in the economy (97/396, 24.5%) or social (79/396, 19.9%) sector. Only a small percentage had prior experience with health training (55/396, 13.9%). Having received psychotherapy in the past was stated by 147 (37.1%) participants, whereas only 35 (8.8%) indicated that they were currently receiving psychotherapy. Table 2 and Table 3 summarize the mean (SD) for the IG and WLC. Concerning the primary outcome at T1, participants had an average value of 22.65 (SD 5.63) on the PSS, thereby indicating a moderate stress level.

Table 1. Baseline characteristics of the study population (N=396).

Characteristics	All participants (N=396)	Intervention group (n=198)	Waiting list control group (n=198)
Sociodemographic characteristics			
Age (years), mean (SD)	41.76 (10.09)	41.96 (10.34)	41.56 (9.87)
Gender, female, n (%)	302 (76.3)	154 (77.8)	148 (74.7)
Married or in a relationship, n (%)	209 (52.8)	106 (53.5)	103 (52)
Having kids, n (%)	185 (46.7)	91 (46)	94 (47.5)
West Germany, n (%)	351 (88.6)	176 (88.9)	175 (88.4)
Ethnicity, n (%)			
Caucasian/White	319 (80.6)	161 (81.3)	158 (79.8)
Asian	9 (2.3)	4 (2)	5 (2.5)
Hispanic	2 (0.5)	2 (1)	0 (0)
Prefer not to say	66 (16.7)	31 (15.7)	35 (17.7)
Education, n (%)			
Low	17 (4.3)	8 (4)	9 (4.5)
Middle	94 (23.7)	52 (26.3)	42 (21.2)
High	285 (72)	138 (69.7)	147 (74.2)
Working characteristics			
Full-time, n (%)	296 (74.7)	144 (72.7)	152 (76.8)
Part-time, n (%)	93 (23.5)	50 (25.3)	43 (21.7)
On sick leave, n (%)	7 (1.8)	4 (2)	3 (1.5)
Management function, n (%)	169 (42.7)	87 (43.9)	82 (41.4)
Work experience years, mean (SD)	17.58 (10.36)	17.62 (10.48)	17.54 (10.28)
Employment status, n (%)			
Permanent	306 (77.3)	158 (79.8)	148 (74.7)
Temporary	42 (10.6)	17 (8.6)	25 (12.6)
Self-employed	35 (8.8)	15 (7.6)	20 (10.1)
Other	13 (3.3)	8 (4)	5 (2.5)
Working sectors, n (%)			
Economy	97 (24.5)	45 (22.7)	52 (26.3)
Service	59 (14.9)	27 (13.6)	32 (16.2)
Social	79 (19.9)	46 (23.2)	33 (16.7)
Health	62 (15.7)	32 (16.2)	30 (15.2)
Information technology	44 (11.1)	21 (10.6)	23 (11.6)
Other	55 (13.9)	27 (13.6)	28 (14.1)
Gross annual income (in €), n (%)			
Low income	89 (22.5)	37 (18.7)	52 (26.3)
Middle income	134 (33.8)	72 (36.4)	62 (31.3)
High income	123 (31.1)	61 (30.8)	62 (31.3)
Not reported	50 (12.6)	28 (14.1)	22 (11.1)
Experience, n (%)			
Previous health training	55 (13.9)	30 (15.2)	25 (12.6)
Previous psychotherapy	147 (37.1)	73 (36.9)	74 (37.4)

Characteristics	All participants (N=396)	Intervention group (n=198)	Waiting list control group (n=198)
Current psychotherapy	35 (8.8)	16 (8.1)	19 (9.6)

Table 2. Means and standard deviations for the intention-to-treat sample at pretreatment.[a]

Outcome	Intervention group (n=198)		Waiting list control group (n=198)	
	Mean (SD)	Range	Mean (SD)	Range
Primary outcome				
PSS[b]	22.39 (5.49)	8-39	22.91 (5.77)	6-37
Mental health				
CES-D[c]	16.02 (7.87)	1-43	16.44 (7.55)	1-38
AQoL8D-MH[d,e]	0.26 (0.11)	0.05-0.61	0.25 (0.12)	0.06-0.56
CD-RISC[e,f]	22.14 (6.63)	1-38	21.29 (7.00)	2-40
MBI-EE-D[g]	4.44 (0.85)	1.8-6	4.43 (0.91)	1.6-6.0
WEMWBS[e,h]	44.36 (7.51)	22-63	42.71 (7.97)	24-62
WHO-5[e,i]	35.09 (16.16)	0-80	35.05 (16.94)	0-96
Work-related outcomes				
ERI-S ratio[j]	1.34 (0.53)	0.42-3.62	1.30 (0.39)	0.29-3.09
OSES[e,k]	23.60 (6.53)	7-36	23.84 (6.35)	6-36
Absenteeism days	5.54 (10.75)	0-66	3.62 (7.97)	0-54
Presentism days	11.63 (13.50)	0-66	12.73 (15.93)	0-66
UWES[e,l]	3.35 (1.18)	0.33-5.89	3.18 (1.27)	0-5.67
WLQ[m] productivity loss	8.44 (5.38)	0-22.82	8.36 (5.14)	0-24.53

[a]Missing data imputed by multiple imputation.
[b]PSS: Perceived Stress Scale.
[c]CES-D: Center for Epidemiological Studies' Depression Scale.
[d]AQoL8D-MH: Assessment Quality of Life 8-Dimensions (mental health component).
[e]Higher scores indicate better outcomes.
[f]CD-RISC: Connor-Davidson Resilience Scale.
[g]MBI-EE-D: Maslach Burnout Inventory (depletion subscale).
[h]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.
[i]WHO-5: 5-item World Health Organization Well-Being Index.
[j]ERI-S: Effort-Reward-Imbalance Questionnaire-Short form.
[k]OSES: Occupational Self-Efficacy Scale.
[l]UWES: Utrecht Work Engagement Scale.
[m]WLQ: Work Limitations Questionnaire.

Table 3. Means and standard deviations for the intention-to-treat sample at posttreatment (T2) and 6-month follow-up (T3).

Outcome	Posttreatment at 7 weeks[a]		At 6-month follow-up[a]	
	Intervention group (n=198), mean (SD)	Waiting list control group (n=198), mean (SD)	Intervention group (n=198), mean (SD)	Waiting list control group (n=198), mean (SD)
Primary outcome				
PSS[b]	16.27 (6.18)	21.22 (6.97)	15.73 (6.07)	19.94 (7.15)
Mental health				
CES-D[c]	11.26 (7.46)	15.52 (8.47)	11.45 (7.21)	15.37 (8.47)
AQoL8D-MH[d,e]	N/A[f]	N/A	0.36 (0.12)	0.29 (0.14)
CD-RISC[e,g]	24.41 (6.72)	20.79 (7.39)	N/A	N/A
MBI-EE-D[h]	3.94 (1.05)	4.31 (0.99)	3.73 (1.05)	4.27 (1.01)
WEMWBS[e,i]	49.28 (7.42)	44.19 (8.34)	50.55 (7.24)	44.02 (8.24)
WHO-5[e,j]	51.25 (18.02)	40.82 (19.92)	54.80 (15.64)	40.79 (20.12)
Work-related outcomes				
ERI-S ratio[k]	1.19 (0.51)	1.24 (0.45)	1.17 (0.40)	1.25 (0.42)
OSES[e,l]	26.21 (5.96)	23.89 (6.76)	N/A	N/A
Absenteeism days	N/A	N/A	4.03 (7.84)	3.22 (6.44)
Presentism days	N/A	N/A	8.88 (9.62)	11.75 (12.91)
UWES[e,m]	3.40 (1.22)	3.02 (1.33)	3.50 (1.03)	3.06 (1.25)
WLQ[n] productivity loss	7.48 (5.67)	8.47 (5.50)	N/A	N/A

[a]Missing data imputed by multiple imputation.
[b]PSS: Perceived Stress Scale.
[c]CES-D: Center for Epidemiological Studies' Depression Scale.
[d]AQoL8D-MH: Assessment Quality of Life 8-Dimensions (mental health component).
[e]Higher scores indicate better outcomes.
[f]N/A: not applicable.
[g]CD-RISC: Connor-Davidson Resilience Scale.
[h]MBI-EE-D: Maslach Burnout Inventory (depletion subscale).
[i]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.
[j]WHO-5: 5-item World Health Organization Well-Being Index.
[k]ERI-S: Effort-Reward-Imbalance Questionnaire-Short form.
[l]OSES: Occupational Self-Efficacy Scale.
[m]UWES: Utrecht Work Engagement Scale.
[n]WLQ: Work Limitations Questionnaire.

Primary and Secondary Outcomes

Tables 4 and 5 display the results of the intention-to-treat analyses for the primary and secondary outcomes. The ANCOVA showed a significant group effect, indicating that participants in the IG had, compared to the WLC, significantly lower scores on the PSS-10 at T2 ($F_{1,393}$=64.44, P<.001) and T3 ($F_{1,393}$=46.91, P<.001) with medium-to-large between-group-effect sizes at T2 (Cohen d=0.71, 95% CI 0.51-0.91) and T3 (d=0.61, 95% CI 0.41-0.81). At posttreatment, significantly more participants in the IG (109/198, 55.1%) showed a reliable improvement on the PSS-10 compared to the WLC (44/198, 22.2%; P<.001). The number needed to treat to achieve 1 additional treatment response (reliable change) at posttreatment was 3.05 (95% CI 2.39-4.20). Significantly less participants experienced a symptom deterioration in the IG (10/198, 5.1%) compared to WLC (19/198, 9.6%), which reflects an absolute risk reduction of 4.55% (95% CI –0.567% to 9.658%) and a relative risk reduction of 47.88% (95% CI –10.307% to 74.888%).

Table 4. Results of the analysis of covariances and Cohen *d* for the primary and secondary outcome measures at posttreatment.

Outcome	At posttreatment after 7 weeks[a] between-groups effect		
	Difference in means (95% CI)	Cohen *d* (95% CI)	ANCOVA[b], *F(1,393)*
Primary outcome			
PSS[c]	–4.66 (–5.80 to –3.51)	0.71 (0.51 to 0.91)	64.44[d]
Mental health			
CES-D[e]	–4.00 (–5.27 to –2.74)	0.55 (0.35 to 0.75)	38.74[d]
AQoL8D-MH[f,g]	N/A[h]	N/A	N/A
CD-RISC[f,i]	2.97 (2.02 to 3.93)	0.54 (0.34 to 0.74)	37.27[d]
MBI-EE-D[j]	–0.38 (–0.54 to –0.23)	0.47 (0.27 to 0.67)	23.60[d]
WEMWBS[f,k]	4.05 (2.81 to 5.29)	0.50 (0.30 to 0.70)	41.31[d]
WHO-5[f,l]	10.41 (7.16 to 13.66)	0.58 (0.38 to 0.78)	39.65[d]
Work-related outcomes			
ERI-S[m] ratio	–0.07 (–0.15 to 0.002)	0.21 (0.01 to 0.41)	3.65
OSES[f,n]	2.49 (1.58 to 3.39)	0.51 (0.31 to 0.71)	29.30[d]
Absenteeism days	N/A	N/A	N/A
Presentism days	N/A	N/A	N/A
UWES[f,o]	0.25 (0.08 to 0.43)	0.23 (0.03 to 0.42)	8.17[d]
WLQ[p]	–1.01 (–2.07 to 0.05)	0.16 (–0.03 to 0.36)	3.51

[a]Missing data imputed by multiple imputation.
[b]Controlling for pretreatment scores (T1).
[c]PSS: Perceived Stress Scale.
[d]Significant at $P<.001$.
[e]CES-D: Center for Epidemiological Studies' Depression Scale.
[f]Higher scores indicate better outcomes.
[g]AQoL8D-MH: Assessment Quality of Life 8-Dimensions (mental health component).
[h]N/A: not applicable.
[i]CD-RISC: Connor-Davidson Resilience Scale.
[j]MBI-EE-D: Maslach Burnout Inventory (depletion subscale).
[k]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.
[l]WHO-5: 5-item World Health Organization Well-Being Index.
[m]ERI-S: Effort-Reward-Imbalance Questionnaire-Short form.
[n]OSES: Occupational Self-Efficacy Scale.
[o]UWES: Utrecht Work Engagement Scale.
[p]WLQ: Work Limitations Questionnaire.

Table 5. Results of the analysis of covariances and Cohen *d* for the primary and secondary outcome measures at posttest and 6-month follow-up.

Outcome	T3[a] Between-groups effect		
	Difference in means (95% CI)	Cohen *d* (95% CI)	ANCOVA[b], *F(1,393)*
Primary outcome			
Perceived Stress Scale	–3.89 (–5.01 to –2.77)	0.61 (0.41 to 0.81)	46.9[c]
Mental health			
Center for Epidemiological Studies' Depression Scale	–3.65 (–4.87 to –2.44)	0.52 (0.32 to 0.72)	34.88[c]
AQoL8D-MH[d,e]	0.06 (0.04 to 0.08)	0.52 (0.31 to 0.72)	32.09[c]
CD-RISC[d,f]	N/A	N/A	N/A
MBI-EE-D[g]	–0.56 (–0.73 to –0.39)	0.61 (0.41 to 0.81)	41.28[c]
WEMWBS[d,h]	5.63 (4.34 to 6.93)	0.66 (0.46 to 0.86)	73.25[c]
WHO-5[d,i]	13.99 (10.78 to 17.21)	0.76 (0.55 to 0.96)	73.32[c]
Work-related outcomes			
ERI-S[j] ratio	–0.10 (–0.17 to –0.03)	0.77 (0.56 to 0.97)	7.53[k]
OSES[d,l]	N/A	N/A	N/A
Absenteeism days	0.48 (–0.91 to 1.87)	0.11 (–0.09 to 0.30)	0.46
Presentism days	–2.44 (–4.40 to –0.49)	0.13 (–0.07 to 0.33)	6.04[m]
UWES[d,n]	0.33 (0.15 to 0.50)	0.25 (0.05 to 0.45)	13.37[c]
WLQ[o]	N/A	N/A	N/A

[a]Missing data imputed by multiple imputation.
[b]Controlling for pretreatment scores (T1).
[c]Significant at *P*<.001.
[d]Higher scores indicate better outcomes.
[e]AQoL8D-MH: Assessment Quality of Life 8-Dimensions (mental health component).
[f]CD-RISC: Connor-Davidson Resilience Scale.
[g]MBI-EE-D: Maslach Burnout Inventory (depletion subscale).
[h]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.
[i]WHO-5: 5-item World Health Organization Well-being Index.
[j]ERI-S: Effort-Reward-Imbalance Questionnaire-Short form.
[k]Significant at *P*<.01.
[l]OSES: Occupational Self-Efficacy Scale.
[m]Significant at *P*<.05.
[n]UWES: Utrecht Work Engagement Scale.
[o]WLQ: Work Limitations Questionnaire.

Comparing reliable changes from baseline to 6 months, 53.5% (106/198) in the IG showed a reliable improvement in contrast to 36.4% (72/198) in the WLC (*P*<.001). The number needed to treat to achieve one additional reliable improvement at 6 months follow-up was 5.82 (95% CI 3.73-13.30). A reliable deterioration was only present in 2% (4/198) of the IG, which was significantly lesser (*P*<.05) than that in the WLC (13/198, 6.6%).

As shown in Tables 4 and 5, the ANCOVAs for mental health outcomes showed significant between-group effects for all outcomes at both assessment points on a *P*<.001 level in favor of the IG. At T2, all effect sizes were moderate ranging from *d*=0.47 for emotional exhaustion to *d*=0.58 for well-being according to the 5-item World Health Organization Well-Being Index. At T3, moderate-to-large effect sizes could be found ranging from *d*=0.52 for mental health to *d*=0.76 for well-being.

Concerning work-related outcomes, the results were less coherent. At T2, the ANCOVAs for work-related self-efficacy (*P*<.001) and work engagement (*P*=.004) showed significant between-group effects, with a small effect size for work engagement (*d*=0.23) and a moderate effect size for occupational self-efficacy (*d*=0.51). There was no significant difference between groups for effort-reward-ratio (*P*=.06) and productivity loss (*P*=.06). However, at T3, apart from absenteeism days

(*P*=.50), all between-group effects were significant (effort-reward-imbalance T3, *P*=.006; presentism days T3, *P*=.01; work engagement T3, *P*<.001), with a large effect size of *d*=0.77 for effort-reward-imbalance and a small effect size of *d*=0.25 for work engagement and presentism days (*d*=0.13).

Adherence

Of the 198 participants in the IG, 188 (95%) finished session 1. Session 2 was completed by 171 (86.4%), session 3 by 154 (77.8%), session 4 by 135 (68.2%), session 5 by 121 (61.1%), session 6 by 108 (54.5%), session 7 by 94 (47.5%), and the booster session by 66 (33.3%) individuals. On average, participants worked through 5.23 (SD 2.74) sessions representing 66% (5.23/8) of the intervention. A linear regression model controlling for baseline stress indicated that the number of completed sessions significantly predicted stress levels at T2 (*b*=–0.67, SE 0.14; *P*<.001; 95% CI –0.94 to –0.39) and T3 (*b*=–0.39, SE 0.14; *P*=.004; 95% CI –0.66 to –0.12). The regression coefficient suggested that the PSS value decreased 0.67 (T2) and respectively 0.39 (T3) points with each additional module completed.

Moderator Analysis

Baseline symptom severity was not found to be a significant moderator, indicating that the intervention can be effective, irrespective of the baseline level of perceived stress. Resilience, agreeableness, psychological strain, and self-regulation competencies significantly moderated the treatment outcome. Resilience moderated the intervention effect at T2 (*P*=.04) as well as at T3 (*P*=.01). The region of significance at T2 and T3 ranged from 1 to 34.28 and 1 to 30.63, respectively, indicating that there was no significant between-group effect for participants with higher levels of resilience at baseline. In total, 2.8% (11/396) of the sample had resilience scores >34.28; 8.8% (35/396) had scores >30.63. Agreeableness moderated the intervention effect on stress at T2 (*P*=.01) with a region of significance from 1 to 4.55, indicating no significant intervention effects within participants with a higher agreeableness score (2 out of 396 participants had agreeableness scores above 4.55). The intervention effect on stress at T3 was moderated by psychological strain (*P*=.04) and self-regulation (*P*=.04). No significant between-group effect was found for participants with a psychological strain score lower than 7.96 (19/396, 4.8%) or self-regulation scores higher than 33.57 (24/396, 6%). Tables S1-S4 in Multimedia Appendix 1 show the results of the moderation analyses whereas Figures 2-6 show a visual representation of the moderation accompanied by a visual representation of the region of significance.

Figure 2. A visual representation of the moderation of the training effect (X) on perceived stress at T2 (Y) by resilience at T1 (M) accompanied with a visual representation of the area of significance according to the Johnson-Neyman technique. IG: intervention group; T1: baseline; T2: at 7 weeks posttreatment; WLC: waiting list control.

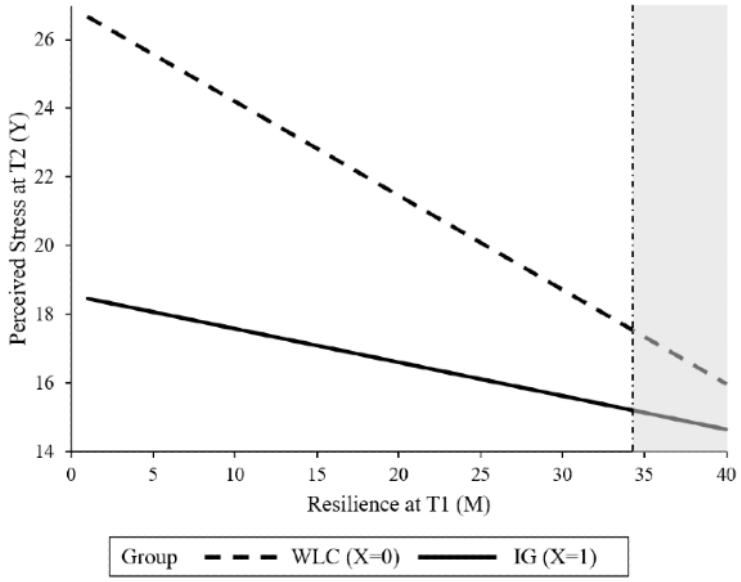


Figure 3. A visual representation of the moderation of the training effect (X) on perceived stress at T3 (Y) by resilience at T1 (M) accompanied with a visual representation of the area of significance according to the Johnson-Neyman technique. IG: intervention group; T1: baseline; T3: at 6 months posttreatment; WLC: waiting list control.

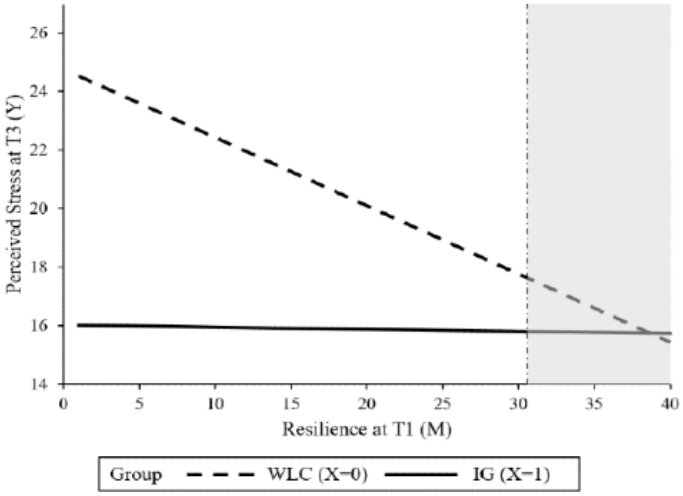


Figure 4. A visual representation of the moderation of the training effect (X) on perceived stress at T2 (Y) by agreeableness at T1 (M) accompanied with a visual representation of the area of significance according to the Johnson-Neyman technique. IG: intervention group; T1: baseline; T2: at 7 weeks posttreatment; WLC: waiting list control.

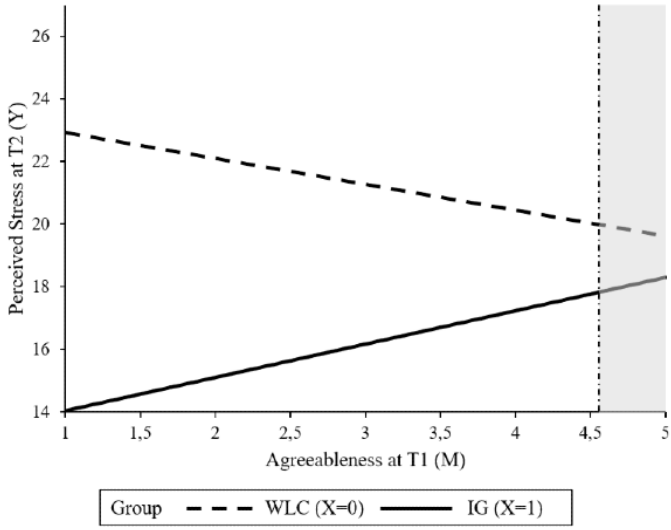


Figure 5. A visual representation of the moderation of the training effect (X) on perceived stress at T3 (Y) by psychological strain at T1 (M) accompanied with a visual representation of the area of significance according to the Johnson-Neyman technique. IG: intervention group; T1: baseline; T3: at 6 months posttreatment; WLC: waiting list control.

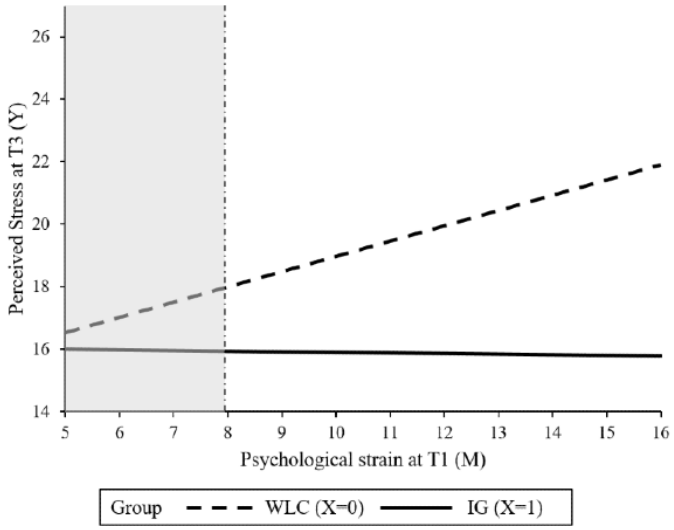
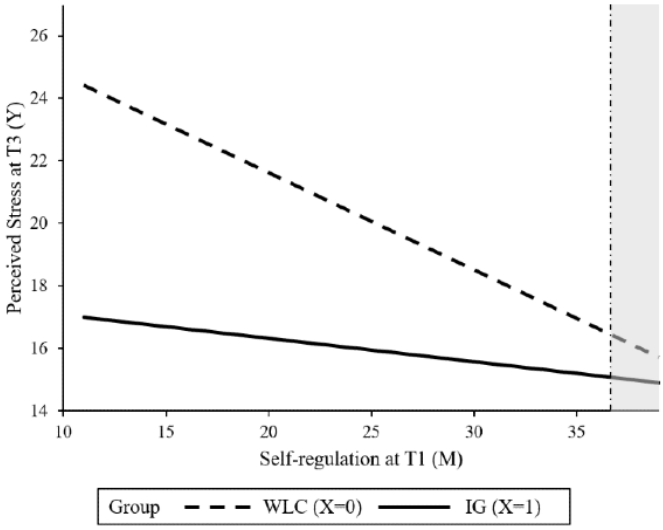


Figure 6. A visual representation of the moderation of the training effect (X) on perceived stress at T3 (Y) by self-regulation at T1 (M) accompanied with a visual representation of the region of significance according to the Johnson-Neyman technique. IG: intervention group; T1: baseline; T3: at 6 months posttreatment; WLC: waiting list control.



No demographic variable (age, gender, relationship status, having kids), other work-related characteristics (full-time work, having a management function, work experience, employment status, working sector, gross annual income), as well as other symptom severity indicators were significantly associated with the intervention effects. *P* values of nonsignificant interaction effects ranged from *P*=.08 (T2) for gender to *P*=.99 for conscientiousness.

Discussion

This study aimed to evaluate the effectiveness and moderators of treatment outcome of a self-guided iSMI as a universal prevention approach without an elaborative inclusion process based on baseline symptom severity. The results of our study indicate that the training was highly effective in reducing stress levels in the short term (*d*=0.71) and long term (*d*=0.61). Significant moderate-to-large effects were found for several secondary mental health and work-related outcomes, including outcomes such as occupational self-efficacy for which no evidence of iSMIs was yet available. Moreover, this study is one of the first to investigate moderator effects in iSMIs in an adequately powered sample. The intervention was suitable for a wide range of participants. Only a few moderators of the intervention effect were identified, indicating that employees with very high resilience, very low psychological strain, very high agreeableness, and very high self-regulation might not profit from the iSMI.

The effects in this study were lower than those in the 4 previously conducted RCTs that examined the same iSMI but in which participants were preselected (PSS>22 or PSS-4 ≥8). With therapeutic guidance, large effect sizes of *d*=0.83 at posttreatment (95% CI 0.58-1.08) were found compared to that of WLC [63], whereas with adherence-focused guidance [23], the between-group effects were *d*=0.79 (95% CI 0.54-1.04), and as a purely self-guided intervention in employees [13] or students [25], the effect size was *d*=0.96 (95% CI 0.70-1.21) or *d*=0.69 (95% CI 0.36-1.02), respectively. However, the CIs in this trial (95% CI 0.51-0.91) overlapped those of all previously conducted studies.

When comparing the results with those from unguided iSMIs not necessarily conducted in the work settings, the effect sizes found in this trial were larger than those found in a recent meta-analysis on the topic (*d*=0.33, 95% CI 0.20-0.46) [70] also when taking 95% CIs into account. Potential reasons for the higher effect sizes compared to other interventions might lie in the strong theoretical basis and the possibility to tailor the intervention to individual need or interest. An alternative explanation might be the strong focus of the intervention on supporting participants to implement health behavior changes in daily life routine. Although the study was powered to test moderation hypotheses adequately, and although we included a broad range of theoretically potential effect modifiers, only few baseline variables significantly moderated the intervention effects. Baseline perceived stress was not associated with intervention outcome, indicating that GET.ON Stress can also be effective in heterogeneous samples that are not preselected based on high baseline symptom severity. That said, it needs to be noted that the sample showed a substantial level of baseline impairment, indicated by a mean of 22.65 (SD 5.63) on the PSS. Despite being lower than that in previous studies that have been conducted on GET.ON Stress (PSS, 25.26-23.90), and although 40.3% (160/396) of the participants in this trial would have been excluded in previous trials, the average baseline symptom severity was still approximately 1 SD above the average perceived stress level in a large working population (mean 15.3 [SD 6.2]) [71]. This might indicate that the intervention might especially be attractive for employees who already experience a substantial stress level. Future studies should hence test whether the utilization of universal preventive approaches in employees can be further increased, for example, by utilizing acceptance-facilitating interventions [72-75] that are designed to reduce the barriers of intervention utilization, such as low perceived risk. Another possibility might be that the format is not adequate for employees with low perceived burden and interventions with even a lower threshold, for example, mobile

apps, focusing on less burdensome behavior changes are necessary to reach this target group.

Resilience, psychological strain, agreeableness, and self-regulation moderated the treatment outcome. Our findings clearly show that the difference in the perceived stress between the IG and the WLC is dependent on the resilience score at baseline. This difference (in the perceived stress between the IG and the WLC) is greater in those that show a lower resilience at baseline, whereas the difference almost becomes nonexistent if the baseline resilience is very high. One potential explanation for this finding might be that very resilient participants manage to cope with perceived stress also in the control group without the help of an additional psychological intervention. Such an interpretation is in line with the theoretical framework around resilience assuming that resilience helps to deal with stressors [76,77]. Several systematic reviews show that high resilience might serve as a protective buffer against the development of psychopathological symptoms [78-82].

Interestingly, employees with low self-regulation profited to a much greater extent from participating compared to those with high self-regulation. This is in contrast to our a priori expectation as we assumed that self-regulation competencies, the ability of an individual to regulate and control their thoughts and behavior enabling them to adapt to a broad range of demands [83], might be a necessary prerequisite for making effective use of such self-help approaches, which require the self-regulated implementation of behavior changes. However, one explanation for the finding might be that participating in the iSMI helps employees with low-self regulation to effectively counteract the missing necessary competencies required to effectively implement health behavior changes in daily life that are needed in order to reduce perceived stress, whereas employees with high self-regulation might manage to realize these necessary self-regulatory tasks also without an additional intervention. Such an assumption of a compensatory effect is supported by the data in the control group showing very high perceived stress level at follow-up in the control group for those with low self-regulation in comparison to low perceived stress levels in the control group for those with high self-regulation. In the IG, the difference between low and high self-regulation is, however, not reflected in the major differences in perceived stress. Nonetheless, empirical studies are needed to confirm these assumptions.

The finding that the intervention was found to be ineffective in individuals with very low current psychological strain (<7.96) has relevant implications for the implementation of such approaches. If these results are confirmed in future studies, such interventions should not be used solely as a universal preventive intervention, for example, to increase protective factors that might help to cope with future stressors. Instead, such interventions should only be offered to employees experiencing at least a minimal level of psychological strain, as this might serve as an important source of motivation that is required for self-help approaches. However, our study was limited to a 6-months follow-up, and future studies are needed to investigate the potential protective effects in such populations regarding future stressors. This study also found agreeableness as a moderator, which has not emerged in the previous literature. Future research should investigate if this was a spurious finding or could lead to a new understanding of the interaction between personality traits and iSMIs.

This study also has some limitations to be considered. First, as usual with RCTs, there was a screening process preceding trial participation, that is, informed consent, which might have caused some applicants with lower motivation to drop out before the study began, potentially resulting in a more homogenous sample. Moreover, the exclusion of suicidal individuals limits the generalizability to such samples. Therefore, although this was a pragmatic trial and eligibility criteria were kept to a minimum to mimic routine care, intervention effects might be overestimated compared to the intervention under routine conditions. Second, one needs to keep in mind that the evidence generated by this study is based on an RCT, which brings a rather high structuring of participants and a high research attention with it, which is usually not the case in routine occupational care. Since the securing of commitment represents an adherence-promoting element in self-help interventions, it can be assumed that the effect sizes for pure self-guided interventions found in RCTs are significantly overestimated for what can be expected in occupational routine care [84] when no additional measures to increase adherence are applied. Hence, a clear concept for ensuring adherence such as through minimal guidance from a professional seems favorable, especially when considering that guided internet-based interventions have found to be superior over pure self-help interventions, both for stress-management [70] as well as other areas [85-87]. Preference should be in routine care, whenever possible, given to self-help approaches with at least some form of adherence-promoting guidance. Moreover, although the study was powered for moderator analyses, we did not stratify for extreme values on assessed moderator variables. Hence, some expressions on the investigated moderators, for example, very low stress levels, were probably not included in sufficient numbers so that a moderator analysis could have detected an existing effect. Further, the power was not sufficient to detect small moderator effects that may nevertheless be of clinical relevance. Since this is one of the first trials on iSMIs as well as SMIs, in general, investigating moderator effects, future empirical studies are needed to confirm our findings.

This study confirms GET.ON Stress to be effective in a heterogeneous sample of employees and to be applicable for a broad range of participant characteristics. Based on the available evidence, iSMIs should be implemented on a broad scale to reduce the adverse consequences of occupational stress.

Acknowledgments

The European Union funded this study (EU EFRE: ZW6-80119999, CCI 2007DE161PR001). BARMER supported the recruitment for this study.

Authors' Contributions

DDE and DL designed the study. DDE, DL, HR, and MB developed the intervention content. DDE and MF performed the outcome analyses. MF prepared the first draft of the manuscript, DDE supervised the writing process, and MF integrated the coauthor comments and edits. All the authors contributed to the further writing of the manuscript and approved the final manuscript.

Conflicts of Interest

DDE and HR are members of the board of the International Society for Research on Internet Interventions. DDE has served as a consultant on the scientific advisory boards of Sanofi, Novartis, Minddistrict, Lantern, Schoen Kliniken, Ideamed, and German health insurance companies (BARMER, Techniker Krankenkasse) and a number of federal chambers for psychotherapy. DDE, MB, BF, and DL are stakeholders in the Institute for health training online (formerly GET.ON, now HelloBetter), which aims to implement scientific findings related to digital health interventions into routine care. HelloBetter distributes the digital intervention under study.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 21 KB - jmir_v23i12e22107_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 660 KB - jmir_v23i12e22107_app2.pdf]

References

1. Stress in America survey. American Psychological Association. 2019. URL: https://www.apa.org/news/press/releases/stress/2019/stress-america-2019.pdf [accessed 2021-11-28]
2. Thoits PA. Stress and health: major findings and policy implications. J Health Soc Behav 2010;51 Suppl:S41-S53. [doi: 10.1177/0022146510383499] [Medline: 20943582]
3. Wiegner L, Hange D, Björkelund C, Ahlborg G. Prevalence of perceived stress and associations to symptoms of exhaustion, depression and anxiety in a working age population seeking primary care--an observational study. BMC Fam Pract 2015 Mar 19;16:38 [FREE Full text] [doi: 10.1186/s12875-015-0252-7] [Medline: 25880219]
4. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012 Dec 15;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
5. Kivimäki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease. Nat Rev Cardiol 2018 Apr;15(4):215-229. [doi: 10.1038/nrcardio.2017.189] [Medline: 29213140]
6. Stansfeld S, Candy B. Psychosocial work environment and mental health--a meta-analytic review. Scand J Work Environ Health 2006 Dec;32(6):443-462 [FREE Full text] [doi: 10.5271/sjweh.1050] [Medline: 17173201]
7. Theorell T, Hammarström A, Aronsson G, Träskman Bendz L, Grape T, Hogstedt C, et al. A systematic review including meta-analysis of work environment and depressive symptoms. BMC Public Health 2015 Aug 01;15:738 [FREE Full text] [doi: 10.1186/s12889-015-1954-4] [Medline: 26232123]
8. Åkerstedt T, Orsini N, Petersen H, Axelsson J, Lekander M, Kecklund G. Predicting sleep quality from stress and prior sleep--a study of day-to-day covariation across six weeks. Sleep Med 2012 Jun;13(6):674-679. [doi: 10.1016/j.sleep.2011.12.013] [Medline: 22621983]
9. Hassard J, Teoh K, Cox T, Cosmar M, Gründler R, Flemming D, et al. Calculating the cost of work-related stress and psychosocial risks: literature review. Publications Office of the European Union. 2014. URL: https://op.europa.eu/en/publication-detail/-/publication/c8328fa1-519b-4f29-aa7b-fd80cffc18cb [accessed 2021-11-28]
10. Bhui KS, Dinos S, Stansfeld SA, White PD. A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. J Environ Public Health 2012;2012:515874 [FREE Full text] [doi: 10.1155/2012/515874] [Medline: 22496705]
11. Richardson KM, Rothstein HR. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
12. Ebert DD, Cuijpers P, Muñoz RF, Baumeister H. Prevention of Mental Health Disorders Using Internet- and Mobile-Based Interventions: A Narrative Review and Recommendations for Future Research. Front Psychiatry 2017;8:116 [FREE Full text] [doi: 10.3389/fpsyt.2017.00116] [Medline: 28848454]
13. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]

14. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
15. Pickett JT, Roche SP. Questionable, Objectionable or Criminal? Public Opinion on Data Fraud and Selective Reporting in Science. Sci Eng Ethics 2018 Feb;24(1):151-171. [doi: 10.1007/s11948-017-9886-2] [Medline: 28281156]
16. Camerer CF, Dreber A, Forsell E, Ho T, Huber J, Johannesson M, et al. Evaluating replicability of laboratory experiments in economics. Science 2016 Mar 25;351(6280):1433-1436. [doi: 10.1126/science.aaf0918] [Medline: 26940865]
17. Camerer CF, Dreber A, Holzmeister F, Ho T, Huber J, Johannesson M, et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat Hum Behav 2018 Sep;2(9):637-644. [doi: 10.1038/s41562-018-0399-z] [Medline: 31346273]
18. Mede NG, Schäfer MS, Ziegler R, Weißkopf M. The "replication crisis" in the public eye: Germans' awareness and perceptions of the (ir)reproducibility of scientific research. Public Underst Sci 2021 Jan;30(1):91-102. [doi: 10.1177/0963662520954370] [Medline: 32924865]
19. Zarski A, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to Internet-Based Mobile-Supported Stress Management: A Pooled Analysis of Individual Participant Data From Three Randomized Controlled Trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
20. Heber E, Ebert DD, Lehr D, Nobis S, Berking M, Riper H. Efficacy and cost-effectiveness of a web-based and mobile stress-management intervention for employees: design of a randomized controlled trial. BMC Public Health 2013 Jul 15;13:655 [FREE Full text] [doi: 10.1186/1471-2458-13-655] [Medline: 23855376]
21. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
22. Ebert DD, Lehr D, Smit F, Zarski A, Riper H, Heber E, et al. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. BMC Public Health 2014 Aug 07;14:807 [FREE Full text] [doi: 10.1186/1471-2458-14-807] [Medline: 25099533]
23. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Sep 01;42(5):382-394 [FREE Full text] [doi: 10.5271/sjweh.3573] [Medline: 27249161]
24. Ebert DD, Kählke F, Buntrock C, Berking M, Smit F, Heber E, et al. A health economic outcome evaluation of an internet-based mobile-supported stress management intervention for employees. Scand J Work Environ Health 2018 Mar 01;44(2):171-182 [FREE Full text] [doi: 10.5271/sjweh.3691] [Medline: 29144535]
25. Harrer M, Adam SH, Fleischmann RJ, Baumeister H, Auerbach R, Bruffaerts R, et al. Effectiveness of an Internet- and App-Based Intervention for College Students With Elevated Stress: Randomized Controlled Trial. J Med Internet Res 2018 Apr 23;20(4):e136 [FREE Full text] [doi: 10.2196/jmir.9293] [Medline: 29685870]
26. 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]
27. Reis D, Lehr D, Heber E, Ebert DD. The German Version of the Perceived Stress Scale (PSS-10): Evaluation of Dimensionality, Validity, and Measurement Invariance With Exploratory and Confirmatory Bifactor Modeling. Assessment 2019 Oct;26(7):1246-1259. [doi: 10.1177/1073191117715731] [Medline: 28627220]
28. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci 2011;13(2):217-224 [FREE Full text] [Medline: 21842619]
29. Alli BO. Fundamental Principles of Occupational Health and Safety. Second Edition. Geneva: International Labor Office; 2008.
30. Lynch FL, Hornbrook M, Clarke GN, Perrin N, Polen MR, O'Connor E, et al. Cost-effectiveness of an intervention to prevent depression in at-risk teens. Arch Gen Psychiatry 2005 Nov;62(11):1241-1248. [doi: 10.1001/archpsyc.62.11.1241] [Medline: 16275811]
31. McLaughlin KA. The public health impact of major depression: a call for interdisciplinary prevention efforts. Prev Sci 2011 Dec;12(4):361-371 [FREE Full text] [doi: 10.1007/s11121-011-0231-8] [Medline: 21732121]
32. Quintiliani L, Sattelmair J, Sorensen G. The workplace as a setting for interventions to improve diet and promote physical activity. 2008 Presented at: WHO/WEF Joint Event on Preventing Noncommunicable Diseases in the Workplace; September 2007; Dalian (China) URL: https://www.who.int/dietphysicalactivity/Quintiliani-workplace-as-setting.pdf?ua=1
33. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693 [FREE Full text] [doi: 10.1017/S0033291716001562] [Medline: 27649340]
34. Junge MN, Lehr D, Bockting CL, Berking M, Riper H, Cuijpers P, et al. For whom are internet-based occupational mental health interventions effective? Moderators of internet-based problem-solving training outcome. Internet Interventions 2015 Mar;2(1):39-47. [doi: 10.1016/j.invent.2014.11.007]

35. 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 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
36. Coudray C, Palmer R, Frazier P. Moderators of the efficacy of a web-based stress management intervention for college students. J Couns Psychol 2019 Nov;66(6):747-754. [doi: 10.1037/cou0000340] [Medline: 30998054]
37. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers PIM, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371. [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
38. Riper H, Hoogendoorn A, Cuijpers P, Karyotaki E, Boumparis N, Mira A, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: An individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714 [FREE Full text] [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
39. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
40. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004 Mar;57(3):229-236. [doi: 10.1016/j.jclinepi.2003.08.009] [Medline: 15066682]
41. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002 Oct;59(10):877-883. [doi: 10.1001/archpsyc.59.10.877] [Medline: 12365874]
42. Beck AT, Steer RA, Brown GK. BDI-II: Beck Depression Inventory Manual. Second edition. San Antonio, TX: Psychological Corporation; 1996.
43. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer Publications; 1984.
44. D'Zurilla TJ, Nezu AM. Problem-solving therapy. In: Handbook of Cognitive-Behavioral Therapies, Third Edition. New York, USA: Guilford Press; 2010:197-225.
45. Berking M. Training Emotionaler Kompetenzen. Berlin, Heidelberg: Springer; 2015.
46. Berking M, Ebert D, Cuijpers P, Hofmann SG. Emotion regulation skills training enhances the efficacy of inpatient cognitive behavioral therapy for major depressive disorder: a randomized controlled trial. Psychother Psychosom 2013;82(4):234-245. [doi: 10.1159/000348448] [Medline: 23712210]
47. Meyer TD, Hautzinger M. Allgemeine Depressions-Skala (ADS). Diagnostica 2001 Oct;47(4):208-215. [doi: 10.1026//0012-1924.47.4.208]
48. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
49. Maxwell A, Özmen M, Iezzi A, Richardson J. Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data. Qual Life Res 2016 Dec;25(12):3209-3219. [doi: 10.1007/s11136-016-1337-z] [Medline: 27344318]
50. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
51. Schaufeli W, Leiter M, Maslach C, Jackson S. Maslach Burnout Inventory-general survey (MBI-GS). American Psychological Association 1996:1-5. [doi: 10.1037/t05190-000]
52. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
53. Bech P, Olsen LR, Kjoller M, Rasmussen NK. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. Int J Methods Psychiatr Res 2003;12(2):85-91 [FREE Full text] [doi: 10.1002/mpr.145] [Medline: 12830302]
54. Li J, Loerbroks A, Jarczok MN, Schöllgen I, Bosch JA, Mauss D, et al. Psychometric properties and differential explanation of a short measure of effort-reward imbalance at work: a study of industrial workers in Germany. Am J Ind Med 2012 Sep;55(9):808-815. [doi: 10.1002/ajim.22018] [Medline: 22298287]
55. Rigotti T, Schyns B, Mohr G. A Short Version of the Occupational Self-Efficacy Scale: Structural and Construct Validity Across Five Countries. Journal of Career Assessment 2008 May 01;16(2):238-255. [doi: 10.1177/1069072707305763]
56. Schaufeli W, Bakker A. UWES-Utrecht Work Engagement Scale Test Manual. Utrecht University: Department of Psychology; 2003. URL: https://www.wilmarschaufeli.nl/publications/Schaufeli/Test%20Manuals/Test_manual_UWES_English.pdf [accessed 2021-11-29]
57. Walker TJ, Tullar JM, Diamond PM, Kohl HW, Amick BC. Validity and Reliability of the 8-Item Work Limitations Questionnaire. J Occup Rehabil 2017 Dec;27(4):576-583 [FREE Full text] [doi: 10.1007/s10926-016-9687-5] [Medline: 28025750]
58. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric

disorder (TiC-P). BMC Health Serv Res 2013 Jun 15;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
59. Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]
60. Schulz H, Lang K, Nübling R, Koch U. Psychometrische Überprüfung einer Kurzform des Fragebogens zur Psychotherapiemotivation - FPTM-23. Diagnostica 2003 Apr;49(2):83-93. [doi: 10.1026//0012-1924.49.2.83]
61. Schwarzer R, Diehl M, Schmitz GS. Self-Regulation Scale. 1999. URL: http://userpage.fu-berlin.de/health/selfreg_e.htm [accessed 2021-11-28]
62. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. American Psychological Association. Windsor, UK: Nfer-Nelson; 1995. URL: https://doi.apa.org/doiLanding?doi=10.1037%2Ft00393-000 [accessed 2021-11-28]
63. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-Based and Mobile Stress Management Intervention for Employees: A Randomized Controlled Trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
64. Leon AC, Heo M. Sample Sizes Required to Detect Interactions between Two Binary Fixed-Effects in a Mixed-Effects Linear Regression Model. Comput Stat Data Anal 2009 Jan 15;53(3):603-608 [FREE Full text] [doi: 10.1016/j.csda.2008.06.010] [Medline: 20084090]
65. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
66. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods 2002;7(2):147-177. [doi: 10.1037/1082-989x.7.2.147]
67. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
68. Hayes AF. Introduction to Mediation, Moderation and Conditional Process Analysis: A Regression-Based Approach. New York, USA: The Guilford Press; 2013.
69. Johnson PO, Fay LC. The Johnson-Neyman technique, its theory and application. Psychometrika 1950 Dec;15(4):349-367. [doi: 10.1007/bf02288864]
70. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
71. Lesage F, Berjot S, Deschamps F. Psychometric properties of the French versions of the Perceived Stress Scale. Int J Occup Med Environ Health 2012 Jun;25(2):178-184 [FREE Full text] [doi: 10.2478/S13382-012-0024-8] [Medline: 22528542]
72. Baumeister H, Nowoczin L, Lin J, Seifferth H, Seufert J, Laubner K, et al. Impact of an acceptance facilitating intervention on diabetes patients' acceptance of Internet-based interventions for depression: a randomized controlled trial. Diabetes Res Clin Pract 2014 Jul;105(1):30-39. [doi: 10.1016/j.diabres.2014.04.031] [Medline: 24862240]
73. Baumeister H, Seifferth H, Lin J, Nowoczin L, Lüking M, Ebert D. Impact of an Acceptance Facilitating Intervention on Patients' Acceptance of Internet-based Pain Interventions. The Clinical Journal of Pain 2015;31(6):528-535. [doi: 10.1097/ajp.0000000000000118]
74. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
75. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A Web-Based Acceptance-Facilitating Intervention for Identifying Patients' Acceptance, Uptake, and Adherence of Internet- and Mobile-Based Pain Interventions: Randomized Controlled Trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925]
76. Pęciłło M. The concept of resilience in OSH management: a review of approaches. Int J Occup Saf Ergon 2016;22(2):291-300 [FREE Full text] [doi: 10.1080/10803548.2015.1126142] [Medline: 26652938]
77. Johnson SJ, Willis SM, Evans J. An examination of stressors, strain, and resilience in academic and non-academic U.K. university job roles. International Journal of Stress Management 2019 May;26(2):162-172. [doi: 10.1037/str0000096]
78. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
79. Leppin AL, Bora PR, Tilburt JC, Gionfriddo MR, Zeballos-Palacios C, Dulohery MM, et al. The efficacy of resiliency training programs: a systematic review and meta-analysis of randomized trials. PLoS One 2014;9(10):e111420 [FREE Full text] [doi: 10.1371/journal.pone.0111420] [Medline: 25347713]
80. Dray J, Bowman J, Campbell E, Freund M, Wolfenden L, Hodder RK, et al. Systematic Review of Universal Resilience-Focused Interventions Targeting Child and Adolescent Mental Health in the School Setting. J Am Acad Child Adolesc Psychiatry 2017 Oct;56(10):813-824 [FREE Full text] [doi: 10.1016/j.jaac.2017.07.780] [Medline: 28942803]

81. Wermelinger Ávila MP, Lucchetti ALG, Lucchetti G. Association between depression and resilience in older adults: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2017 Mar;32(3):237-246. [doi: 10.1002/gps.4619] [Medline: 27805730]
82. Brassington K, Lomas T. Can resilience training improve well-being for people in high-risk occupations? A systematic review through a multidimensional lens. The Journal of Positive Psychology 2020 Apr 27;16(5):573-592. [doi: 10.1080/17439760.2020.1752783]
83. Baumeister RF, Vohs KD. Self-Regulation, Ego Depletion, and Motivation. Social Pers Psych Compass 2007 Nov;1(1):115-128. [doi: 10.1111/j.1751-9004.2007.00001.x]
84. Ebert DD, Baumeister H. Internet-Based Self-help Interventions for Depression in Routine Care. JAMA Psychiatry 2017 Aug 01;74(8):852-853. [doi: 10.1001/jamapsychiatry.2017.1394] [Medline: 28636717]
85. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
86. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for panic disorder: A three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. J Anxiety Disord 2018 May;56:43-55. [doi: 10.1016/j.janxdis.2018.03.009] [Medline: 29625834]
87. Lin J, Paganini S, Sander L, Lüking M, Ebert DD, Buhrman M, et al. An Internet-Based Intervention for Chronic Pain. Dtsch Arztebl Int 2017 Oct 13;114(41):681-688 [FREE Full text] [doi: 10.3238/arztebl.2017.0681] [Medline: 29082858]

Abbreviations

ANCOVA: analysis of covariance
IG: intervention group
iSMI: internet-based mobile-supported stress management intervention
PSS: perceived stress scale
RCT: randomized controlled trial
WLC: waiting list control

Edited by G Eysenbach; submitted 03.07.20; peer-reviewed by L Genaro, M Adly, A Adly, A Adly; comments to author 24.08.20; revised version received 15.12.20; accepted 17.05.21; published 22.12.21.

Please cite as:
Ebert DD, Franke M, Zarski AC, Berking M, Riper H, Cuijpers P, Funk B, Lehr D
Effectiveness and Moderators of an Internet-Based Mobile-Supported Stress Management Intervention as a Universal Prevention Approach: Randomized Controlled Trial
J Med Internet Res 2021;23(12):e22107
URL: https://www.jmir.org/2021/12/e22107
doi:10.2196/22107
PMID:34941541

Original Paper

Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study

Yuepei Xu[1,2*], BSc; Ling-Zi Yue[1,2*], MSc; Wei Wang[3], PhD; Xiao-Ju Wu[1,2], MSc; Zhu-Yuan Liang[1,2], PhD

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China

[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[3]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Zhu-Yuan Liang, PhD
CAS Key Laboratory of Behavioral Science
Institute of Psychology
16 Lincui Road, Chaoyang District
Beijing, 100101
China
Phone: 86 10 64841536
Email: liangzy@psych.ac.cn

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/3/e38221

Abstract

Background: Walking is a simple but beneficial form of physical activity (PA). Self-monitoring and providing information about social norms are the 2 most widely used "mobile health (mHealth)" strategies to promote walking behavior. However, previous studies have failed to discriminate the effect of self-monitoring from the combination of the 2 strategies, and provide practical evidence within Chinese culture. Some essential moderators, such as gender and group identity, were also overlooked.

Objective: We aimed to investigate the effectiveness of social norm and self-monitoring interventions for walking behavior and assess the moderating effects of gender and group identity, which could guide optimal mHealth intervention projects in China.

Methods: In 2 longitudinal tracking studies (study 1, 22 days; study 2, 31 days), Chinese college students wore trackers for at least 8 hours per day (MASAI 3D Pedometer and Xiaomi Wristband 2) to record their daily step counts in baseline, intervention, and follow-up stages. In each study, participants (study 1: n=117, 54% female, mean age 25.60 years; study 2: n=180, 51% female, mean age 22.60 years) were randomly allocated to 1 of the following 3 groups: a self-monitoring group and 2 social norm intervention groups. In the 2 intervention groups and during the intervention stage, participants received different social norm information regarding group member step rankings corresponding to their grouping type of social norm information. In study 1, participants were grouped by within-group member PA levels (PA consistent vs PA inconsistent), and in study 2, participants were grouped by their received gender-specific social norm information (gender consistent vs gender inconsistent). Piece-wise linear mixed models were used to compare the difference in walking steps between groups.

Results: In study 1, for males in the self-monitoring group, walking steps significantly decreased from the baseline stage to the intervention stage (change in slope=−1422.16; *P*=.02). However, additional social norm information regardless of group consistency kept their walking unchanged. For females, social norm information did not provide any extra benefit beyond self-monitoring. Females exposed to PA-inconsistent social norm information even walked less (slope during the intervention=−122.18; *P*=.03). In study 2, for males, a similar pattern was observed, with a decrease in walking steps in the self-monitoring group (change in slope=−151.33; *P*=.08), but there was no decrease in the 2 social norm intervention groups. However, for females, gender-consistent social norm information decreased walking steps (slope during the intervention=−143.68; *P*=.03).

Conclusions: Both gender and group identity moderated the effect of social norm information on walking. Among females, social norm information showed no benefit for walking behavior and may have exerted a backfire effect. Among males, while walking behavior decreased with self-monitoring only, the inclusion of social norm information held the level of walking behavior steady.

(J Med Internet Res 2021;23(12):e29167) doi:10.2196/29167

KEYWORDS

self-monitoring; social norm; group identity; gender differences; mHealth; mobile health

Introduction

Background

Walking is a simple but highly beneficial form of physical activity (PA) [1]. It can lower the risk of premature death [2] and prevent numerous chronic diseases such as obesity, type II diabetes, and cardiovascular disease [3]. In addition to providing reasonably accurate daily walking records [4], smartphones are also capable of additional functions such as tracking the number of steps walked. Such functionality has led to the use of mobile devices to promote a healthier lifestyle. Indeed, this concept of a mobile health (mHealth) strategy has become an increasingly popular public health intervention tool to promote walking behavior [5].

Providing social norm information and providing one's self-monitoring information (ie, a record of one's behavior) are currently 2 of the most widely used strategies in mHealth. In self-monitoring interventions, pedometers or pedometer apps on smartphones (eg, Apple Health and Accupedo) provide users with daily step counts, which enable them to monitor and improve their own walking behavior [5-8]. Combined with this self-monitoring strategy, several leading online social media platforms (eg, Facebook and WeChat) have also implemented a social norm strategy to form an mHealth intervention for walking behavior [9,10]. The biggest difference between the 2 strategies is that the latter strategy allows users to view their own records and the walking records of their peers. By exposing users to a highly social norm sensitive context, users might be impacted by the perceived levels of others' PA and might thereafter increase their own walking behavior [11-13]. Previous studies examining the effect of social norms have shown that there is still some debate regarding its veracity [14,15], and others have failed to consider the influence of some potential moderators [16]. Thus, it is still unclear as to whether a social norm strategy would be effective in promoting walking behavior.

In this study, we aimed to optimize mHealth walking interventions and provide more practical evidence about the effectiveness of interventions based on social norms. To address these aims, we implemented 2 longitudinal tracking studies in order to compare and isolate the mixed effects of social norm and self-monitoring interventions on promoting walking. In particular, we focused on 2 theoretically essential factors and ignored moderators regarding the influence of social norm interventions (group identity and gender).

Self-Monitoring Intervention: Effect of One's Own Walking Records

Self-monitoring refers to a systematic self-observation or recording of target behaviors (eg, walking and food intake) [17]. Based on the self-regulation theory, self-monitoring is an essential component of a successful self-regulation process, preceding self-evaluation and self-reinforcement [18]. Self-monitoring thus motivates individuals to make changes by stimulating them to focus, evaluate, and regulate the target behavior [17]. Several studies have shown that the strategy of self-monitoring plays a positive role in promoting a variety of healthy or healthier behaviors such as reduction in smoking [19] and improvements in diet [20].

Data from several studies suggest that a self-monitoring strategy, typically informing people how many steps they have walked, can be effective in boosting walking behavior [6-8]. For example, one study showed that by wearing a pedometer and setting a self-monitored goal of walking 10,000 steps each day, overweight women were able to increase their average walking steps by 85% (from 4972 to 9213 steps per day) [21]. A recent meta-analysis based on 6 interventions from 2005 to 2017 also showed that a self-monitoring strategy was associated with an increase in walking of up to 3090 steps per day in cardiovascular patients [22]. In a review of 88 intervention studies (N=18,804), Knittle et al concluded that self-monitoring is one of the most widely used strategies and represents an essential component for increasing both the intention of PA (b=0.30) and actual PA behavior (b=0.28) [23]. As a consequence, self-monitoring has generally been adopted as the centerpiece of most public health intervention programs [24,25].

However, there has been some controversy surrounding the effect of self-monitoring. For example, some researchers have argued that, despite its numerical effectiveness, the act of self-monitoring does not have a positive influence on walking behavior. Self-monitoring may also not guarantee high levels of exercise adherence, with rates of adherence shown to decline significantly over time [26].

More importantly, previous studies have paid little attention to exploring the mechanisms that underlie self-monitoring interventions in practice, thus limiting its application value and generalizability to other areas. For example, in order to optimize effectiveness, interventions have implemented a combination of strategies with self-monitoring included in a package of self-regulation processes [17], such as goal setting [21]. As a consequence, it is difficult to disambiguate the actual effect of self-monitoring from the mix of other potential effects. In addition, the majority of previous intervention studies have

failed to consider the potential for certain demographic variables to exert a moderating effect on the relationship between self-monitoring and PA behavior [17]. Gender represents a key example whereby females may benefit more from a self-monitoring intervention and maintain an active lifestyle for longer than males [27].

Social Norm Intervention: Effect of the Walking Information of Peers

Numerous theories of behavioral change, including the theory of planned behavior as an example, have listed social norms as an essential predictor of actual changes in behavior [28]. Previous studies have demonstrated that 2 types of social norms [29,30] are important for behavior change across areas ranging from reducing alcohol use [15,31] to reducing environmental damage [30]. These are descriptive norms, which refer to the perceived prevalence of target behaviors, and injunctive norms, which refer to perceived approval or disapproval of target behaviors by the society. Descriptive social norms may guide the behavior of individuals because of implied social proof that the behavior of the majority should be right, as the focus theory of normative conduct posits [30]. Injunctive norms may guide behavior because of a need by individuals to obtain social approval for meaningful relationships with other ingroup members [29,30].

In particular, the perception of social norms about PA was positively correlated with the actual PA level of individuals [13,32], and providing social norm information can even lead individuals to do more PA [12]. For example, compared with traditional online tracking or sending of promotional messages, interventions based on social norms, such as sharing the information of peers, led to participants spending more time on PA [10]. The typical use of descriptive norms, such as sending a message containing exercise information of colleagues, was more effective in increasing mild physical exercise than interventions that did not have any norm manipulation, such as sending general information about physical exercise and health [12]. McEachan et al analyzed 100 empirical studies from 1990 to 2010 (N=22,849) and found social norms to be a strong predictor for PA (mean correlation ρ=0.21, 95% CI 0.18-0.24) [33]. A series of meta-analyses [28,34] also revealed that changing one's perception of social norms can produce a small to moderate effect (Cohen *d*=0.36) in increasing healthy behavior, including physical exercise.

However, the notion that social norm interventions are reliably effective has been complicated by some studies that have found its effect on behavioral change to be unstable or even negative [14,15]. For example, a meta-analytic review reported only a weak correlation (*r*=0.17, 90% CI −0.07 to 0.43) between social norms and PA [35]. The relationship between social norms and exercise intention may disappear after controlling for other relevant factors such as attitude to sports [36]. Indeed, a situation in which a desirable behavior is only a minority norm (eg, only a few people walk over 10,000 steps every day) or in which an undesirable behavior is the majority norm (eg, the majority only walk 2000 steps every day) may even lead to a backfire effect where an individual's healthy behavior is reduced [37,38], suggesting that social norm interventions must be applied with caution. Even from a methodological level, since social norm interventions are generally combined with self-monitoring (ie, users may track their own walking and their peers' walking at the same time [39]), it can become difficult to detect the pure effect of social norm interventions. Consequently, given that the implied factors that can lead to success or failure in these social norm intervention studies are unclear, researchers have argued that the likelihood of both type 1 and type 2 errors is high [38]. This has led some to suggest that social norm interventions should be dropped from further use [36]. Moreover, compared with correlational studies measuring subjective social norms, evidence-based interventions targeting social norms are still limited [36].

Group Identity and Gender as Moderators of the Effects of Social Norm Interventions on PA

With the development of social norm theories, researchers have increasingly declared that the relationship between social norms and target behavior cannot be explained by a simple one-way causal model. Recent studies have attempted to explain the weak effect of social norm interventions on PA by investigating previously ignored moderators such as group identity [16]. However, most previous interventions simply provided social norm information by sending a message without investigating any other potential variables [10,12]. This may lead to a disparity between an up-to-date theory framework and previous practical interventions.

As one of the most widely used and effective theories in predicting and intervening in health behaviors, such as reducing alcohol intake [40] and promoting hand-washing [41], the theory of normative social behavior (TNSB) has proposed group identity as a factor moderating the effect of descriptive norms on behavior [42]. Group identity is defined as the degree to which an individual perceives similarity with the group and aspires to emulate group members [42]. The TNSB predicts that the effect of social norm interventions on behavioral change would be stronger when individuals perceive a higher similarity between themselves and the reference group [42], that is, social norms from a closer group work better at changing behaviors due to higher group identity, and individuals are more willing to perform the target behavior as an expression of group solidarity [42]. This moderating effect of group identity on the effect of social norm interventions can also be explained by the phenomenon of groupthink. Groupthink is more likely to take place in groups with higher identity and may put pressure on group members to conform to the social norms of the group [43]. Empirical studies support the prediction of the TNSB. For example, the closer individuals felt to a group, the closer their own behavior was to their perception of the social norms of that group [44]. Social norm interventions have been found to promote office colleagues with a high group identity to play more sports, while the effect was insignificant among college students who had a low group identity [12].

As an innate label, information regarding individuals with the same gender may feel more relevant and relatable. Therefore, according to the TNSB, because of a higher group identity, social norms within males or females (gender-consistent or gender-specific social norms) may have a greater impact on the

behaviors of individuals [45]. For example, one study showed that perceived gender-consistent norms were stronger predictors of alcohol consumption than gender-inconsistent norms, and this was especially evident for females [46]. However, although previous studies have suggested that manipulating gender-consistent norms may be more beneficial in norm-based interventions [45,46], the availability of practical evidence is still limited and controversial. For instance, an intervention using gender-consistent norms did not show a better effect on reducing alcohol consumption than gender-inconsistent norms [47]. Gender may also interact with the effect of social norm interventions on target behaviors (different genders may have divergent reactions after receiving the same normative information). For instance, a social norm intervention regarding driving behavior may be effective for males but not females because of stronger normative pressure regarding this particular behavior in men compared to women [48]. On the other hand, females may be more sensitive to norms related to their body appearance than males due to their increased tendency to engage in appearance-related social comparison [49]. Even though these studies clearly suggest that the effectiveness of social norm interventions is strongly associated with gender, the importance of this factor has been largely overlooked in previous studies investigating mHealth interventions.

Moreover, the finding of systematic gender differences in the target behavior of this study (ie, PA) (see Pollard and Wagnild's review [50]) calls for the need to consider gender as a moderating factor for the effect of social norm interventions on walking behavior. For example, walking has been observed to be a type of PA preferred by females [51], although consistent evidence has shown that males did more physical exercise than females [51-53]. According to evidence showing that self-regulation was the best predictor of PA in female college students but not males, females may be more sensitive to self-monitoring interventions [52]. Given that China currently has one of the widest gender gaps in the world [54], these gender differences within the Chinese cultural context may be more pronounced than those within other regions. In contrast to the commonly accepted gender-consistent attitude to sports, in China, males reported a more positive attitude to sports than females [55]. Compared with female college students, male students were also found to exercise more frequently, and they also showed both a higher participation rate and voluntary motivation to practice sports [56].

Our Study

Since 2018, China has been the home of over 500 million mHealth users [57], making it an ideal region to evaluate the effects of different interventions. The use of smartphone pedometer apps has been found to be more favorable to traditional wearable pedometer devices in improving walking behavior [58]. Consequently, some social media platforms began to design their own pedometer plugins. WeChat, one of the most popular social media platforms with over 1 billion users in China [59], designed WeRun to provide daily step tracking and step ranking among the contacts of WeChat users. In addition to viewing rankings, users can give a thumbs-up or follow the step records of their contacts. With 15.3% of users taking advantage of its novel social functions, WeRun has become one of the most widely used features of WeChat since its launch in 2015. According to WeRun users, they use it to not only get data about their own exercise behavior, but also increase exercise frequency as well as interaction with friends [60]. As a consequence, WeChat represents a highly appropriate platform for investigating the effect of social norm interventions.

Although there are increasing numbers of apps, smart devices, and high-tech companies trying to promote PA among individuals through systems that utilize self-monitoring or social norms, little is known about the comparison of the effects of social norm interventions with self-monitoring interventions. The absence of substantial research investigating the underlying mechanisms of social norm interventions has obstructed the development of more personalized and optimized mHealth interventions. This disconnection between up-to-date social norm theories (eg, the TNSB) and antiquated intervention techniques may have weakened the potential for practical benefits from the existing evidence base.

Moreover, previous studies have suffered from some limitations, both methodologically and practically. For example, since most previous studies on social norms lacked a corresponding self-monitoring control group [39], the mixed-strategy approaches may have not only resulted in ambiguous comparisons of the effects of social norm and self-monitoring information on walking behavior, but also led to the collection of data characterized by low reliability and high error rates [38]. In addition, given that most mHealth interventions are specific to a certain app or device, the particular results of the corresponding interventions were device-specific and were thus not stable enough to be generalizable. In the context of mHealth, China is a rapidly developing country, but it is also suffering from a specific social issue in the shape of a wide gender gap. We have very limited understanding as to how gender and different mHealth interventions interact, and thus, there is no useful guidance as to how mHealth interventions in China can be optimized.

To address this, we aimed to answer the following 2 primary questions in this study: (1) In China, is an intervention based on social norms more effective than self-monitoring at promoting walking behavior? (2) Do gender and group identity moderate the effect of social norm interventions on walking? Specifically, we conducted 2 longitudinal tracking studies to compare the effects of self-monitoring and social norms on walking behavior. In study 1, we compared the effects of social norm and self-monitoring interventions on PA between genders by randomly assigning participants to a self-monitoring group, a PA-consistent intervention group (ie, similar levels of PA among group members), or a PA-inconsistent group (ie, varying levels of PA among group members). In study 2, after observing a gender-specific effect of social norm interventions on PA, we attempted to replicate the main finding of study 1 using a more precise measure of steps walked. We also assigned participants to a self-monitoring group, a gender-consistent intervention group (ie, providing social norm information that is gender-specific), or a gender-inconsistent intervention group (ie, providing social norm information that is not gender-specific). Our hypotheses were as follows: hypothesis 1 (H1), in China, providing social norm information, in the form

of step ranking, can promote walking more effectively compared with self-monitoring; hypothesis 2 (H2), gender will moderate the effect of social norm interventions on the promotion of walking, and the walking behavior of males will be more impacted by social norms; and hypothesis 3 (H3), group identity will moderate the effect of social norm interventions on the promotion of walking, and a higher sense of group identity will strengthen the effect of social norm interventions. Furthermore, H3 had the following 2 parts: H3a, an intervention using PA-consistent norms will have a larger effect on walking behavior than that using PA-inconsistent norms; H3b, an intervention using gender-consistent norms will have a larger effect on walking behavior than that using gender-inconsistent norms.

Methods

Participants

In both studies, we recruited graduate students from the University of Chinese Academy of Sciences, Beijing, China as our participants by sending advertisements during university courses (744 students in study 1, and 986 students in study 2). To avoid the possible confounding effect of walking habits or previous mHealth app experience, all recruited students completed a 5-minute online screening and customized sports habit questionnaire. The inclusion criteria were as follows: (1) used WeRun or other walk-related mHealth apps less than twice a week; (2) not a member of any WeChat sports group; (3) self-reported being mentally and physically healthy; (4) consented to participate in the entirety of the experiment. Next, according to their self-reported sports habits, participants were labeled as either low PA (sports fewer than twice a week and run a total distance of less than 5 km per week) or high PA (sports more than three times per week and run a total distance of more than 10 km per week).

Given that the motivation of this study was to promote walking behavior among the low-PA sample, we only selected low-PA students as our formal participants (high-PA students were only recruited in study 1 to set up the PA-inconsistent group, but were not included for further analysis). We preset our sample size to at least 40 participants in each group. Thus, we included a total of 127 low-PA participants in study 1 (with another 42 high-PA students), while 182 low-PA participants were recruited for study 2.

This study was approved by the Institutional Review Board of the Institution of Psychology, Chinese Academy of Sciences, and all participants took part in the study voluntarily and completed written informed consent. To record the number of steps walked, all participants were asked to wear step trackers (MASAI 3D Pedometer, study 1) or smart bands (Xiaomi Wristband 2, study 2) during the entire experimental period. After the study, participants were allowed to keep their step trackers (and approximately US $18 in study 1) or smart bands as payments for participation.

Design and Procedures

Study Design

To compare the effects of self-monitoring and social norms, both studies consisted of 3 stages (baseline, intervention, and follow-up; Figure 1 and Figure 2), with slight variations in the length of each stage. Study 1 was conducted from April to May 2017 and consisted of 22 days, with 3, 14, and 5 days in each stage, respectively. Study 2 was conducted from October to November 2017 and consisted of 31 days, with 10, 15, and 6 days in each stage, respectively. All participants were required to wear the trackers during the entire experiment and were able to check their step counts at any time in order to monitor their own walking behavior.

Figure 1. Flow diagram for study 1. The diagram shows the complete experimental procedure including enrollment, randomization, and intervention. All participants (valid n=117) in 3 groups were tracked for 22 days. PA: physical activity.

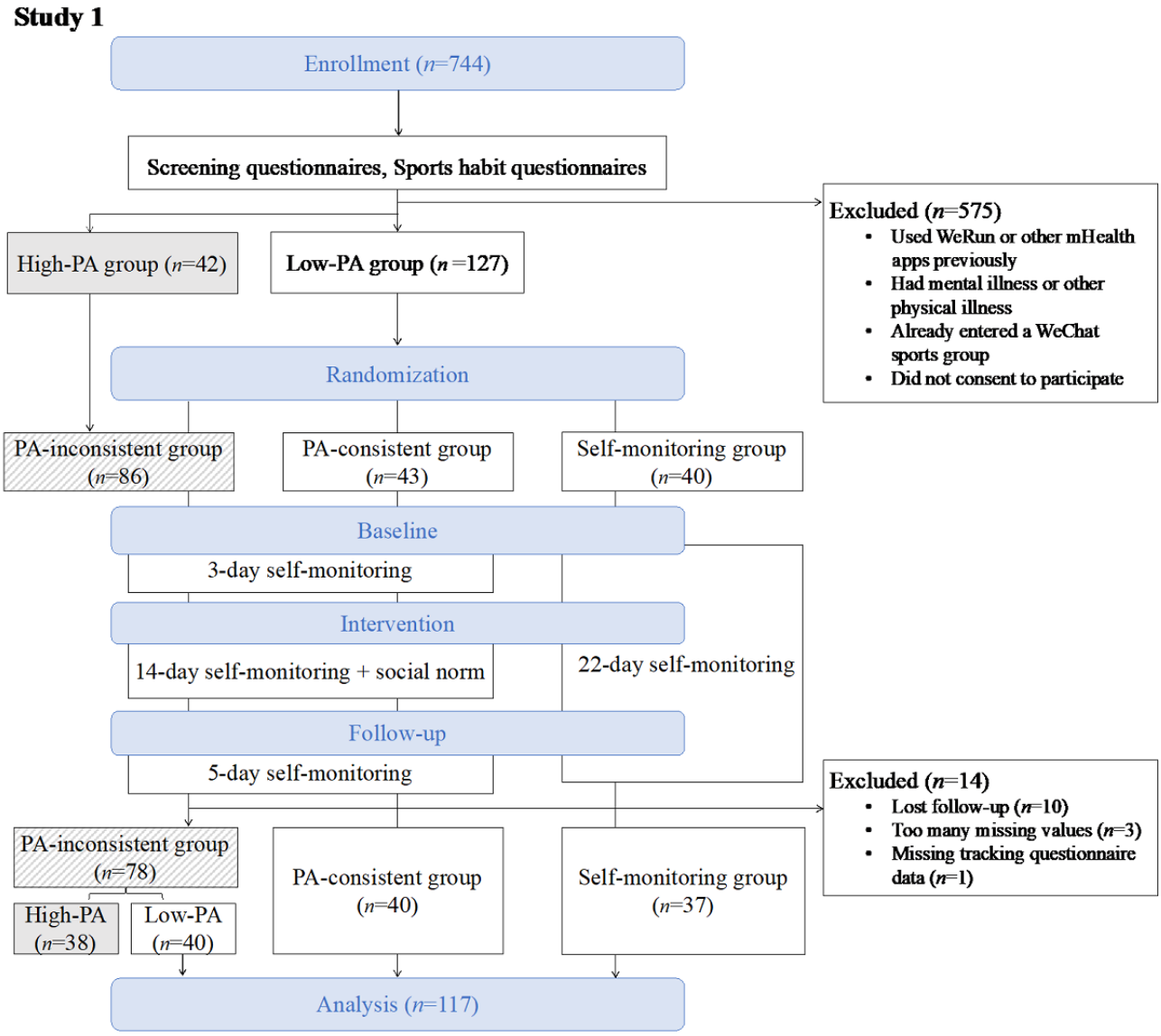
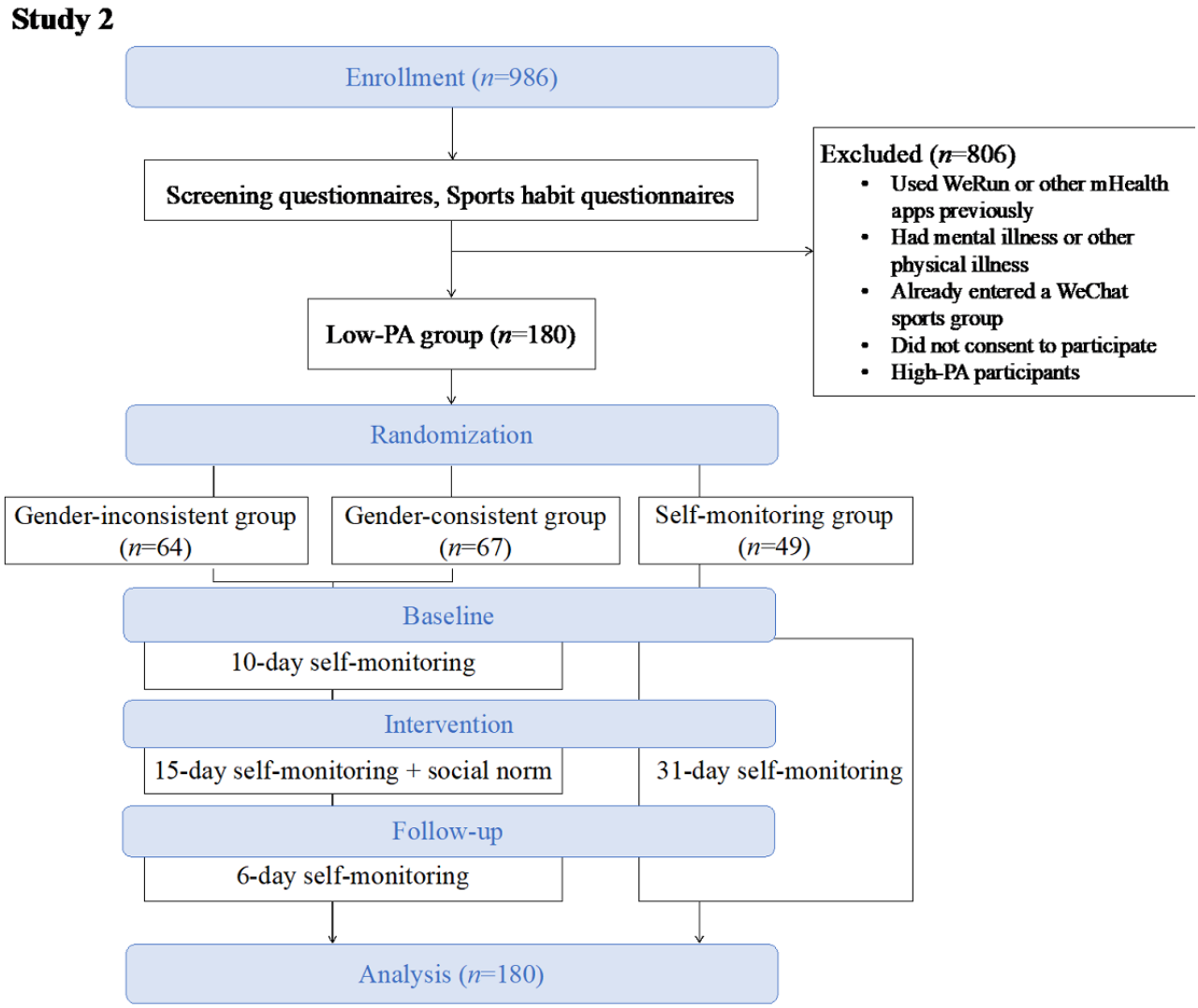


Figure 2. Flow diagram for study 2. The diagram shows the complete experimental procedure including enrollment, randomization, and intervention. All participants (valid n=180) in 3 groups were tracked for 31 days. PA: physical activity.



To compare the effects of self-monitoring and social norms, we randomized all participants in both studies into 3 groups according to the interventions they received, namely, a self-monitoring intervention group and 2 social norm intervention groups. Participants in the 2 social norm intervention groups received social norm information from experimenters during the intervention stage, whereas participants in the self-monitoring intervention group did not receive any extra message during the experimental stages.

To test the moderating effect of group identity on the effect of social norm interventions, we aimed to compare the effects of providing social norm information among groups that should theoretically have a stronger sense of group identity (ie, PA-consistent and gender-consistent groups) against those that should theoretically have a weaker sense of group identity (ie, PA-inconsistent and gender-inconsistent groups). In particular, in study 1, we randomly assigned participants to either the self-monitoring group or 2 social norm intervention groups (the PA-consistent intervention group and PA-inconsistent intervention group). In the PA-consistent intervention group, all participants had a low PA level, while in the PA-inconsistent intervention group, half of the participants had a low PA level and half had a high PA level. In study 2, we randomly assigned participants to either a self-monitoring group or 2 social norm intervention groups (the gender-consistent intervention group and gender-inconsistent intervention group). In the gender-inconsistent (but PA-consistent) intervention group, participants received general gender-mixed social norm information as in study 1 (ie, gender-mixed information simply showing the rank of each member in the WeChat group), while in the gender-consistent (and PA-consistent) intervention group, participants received 2 gender-separate social norm information sets (ie, 2 gender-specific information sets showing the rank of each female student among all the female members and each male student among all the male members). Through this design, participants in the PA-consistent group in study 1 and gender-consistent group in study 2 will theoretically have a stronger sense of group identity than participants in the other groups.

Except for high-PA participants in study 1, all low-PA participants in both studies were randomly allocated to 1 of the 3 groups according to randomly generated numbers. All participants were single-blind to the nature of our experimental design during the entire experimental period.

Intervention Procedures

During the whole experimental period (including baseline, intervention, and follow-up), each participant was able to monitor his/her own steps by checking the trackers at any time. In addition, in order to send social norm information within the 2 social norm intervention groups of both studies, during the intervention stage, we set up separate WeChat groups for participants within each intervention group. Participants in the WeChat groups were asked to send a picture of their step trackers (study 1) or smart bands (study 2) before sleep every day. The next day at 10 AM, experimenters sent a personalized step ranking to each WeChat group member regarding their performance from the previous day.

To ensure that participants received the daily step ranking and paid sufficient attention to it, they were asked to finish an online daily tracking questionnaire containing 3 questions about step rankings (eg, How many steps did the person rank first in the group walk yesterday?) and their step tracker wearing habits (eg, When did you start to wear the tracker today?).

Social Norms and Other Measurements

To measure social norms as our secondary outcomes, participants also completed a social norm questionnaire on the last day of each stage (total of 3 times) that was adapted from a previous study [61]. The injunctive social norm questionnaire (6 items) asked participants to judge on a 7-point Likert scale (1, not agree at all to 7, totally agree) whether they agreed on some positive statements about playing sports (eg, most successful people have habits associated with playing sports). The descriptive gender norm questionnaire (2 items) measured participants' descriptive male gender norms on walking (estimation of the percentage of male students in the same university walking over 6000 steps) and descriptive female gender norms on walking (estimation of the percentage of female students in the same university walking over 6000 steps). Moreover, participants were asked to self-report some demographic information, including their age, subjective health (on a 5-point Likert scale; 1, very unhealthy to 5, very healthy), height (m), and weight (kg), and BMI was calculated (weight/height^2).

Data Cleaning

Consistent with previous studies [39], we deleted data points for the following reasons: participants failed to answer all the daily tracking questions, device wear time was shorter than 8 hours (in study 1), or step count was less than 1000. In study 1, 4.72% of data points of step counts were missing, and in study 2, 0.79% of data were missing. Days with nonmissing data from the same participant were still included in the final analyses.

In study 1, we excluded a total of 14 participants from all analyses for the following reasons: 10 were lost in the follow-up stage, 1 did not answer the tracking questionnaire during the intervention stage, 1 had no valid baseline step data, and 2 lacked sufficient intervention step data (more than five missing values). The final valid sample included 155 participants (75 males and 80 females; mean age 25.75 years, SD 1.27 years; age range 23-33 years), of which 117 participants were from low-PA groups. In study 2, 2 participants quit the experiment halfway through and were thus excluded from all analyses. This left a valid sample of 180 participants (92 males and 88 females; mean age 22.60 years, SD 1.16 years; age range 20-30 years).

Statistical Methods and Data Analysis

Baseline Comparison

In order to ensure the homogeneity of all the experimental groups, we conducted analysis of variance (ANOVA) to compare participant characteristics (age, subjective health level, and BMI) and baseline average step counts between the 3 groups. The homogeneity of variance was tested and satisfied. The significant factor (BMI in study 1) was included as a covariate in the analyses for the main outcomes.

Analysis for Social Norms

In order to compare changes in social norms (injunctive norms, male descriptive norms, and female descriptive norms) among groups across different experiment stages, we used a linear mixed-effects (LM) model with the stage (baseline, intervention, and follow-up) as a within-group factor, group and gender (male and female) as between-group factors, BMI in study 1 as a covariate, and the participant as a random intercept term in the model.

Analysis of Walking Step Data

Walking step data in our studies had 2 key features. First, there were 3 different stages (baseline, intervention, and follow-up) in the entire experiment. Second, each stage consisted of a segment of repeatedly measured daily step counts from each participant (in study 1, a total of 22 daily step counts, while in study 2, a total of 31 daily step counts). Given these features, to better capture the variability of participants' daily step counts, both at the individual level and the trend during each stage, we applied a piece-wise linear mixed-effects (PLM) model. Compared with a traditional simple linear model with a constant slope, a piece-wise linear model implements several segmented regression lines with different slopes, and thus is suitable to perform staged and especially time-segmented prediction [62]. Meanwhile, the mixed effects of the PLM take into account the individual differences during repeated measurements in each experiment stage [63].

The PLM model for walking step data was conducted for each gender separately. The predictors of the model were group, time, and their interactions, with BMI as a covariate in study 1 and the participant as a random intercept term in the model. The predictor "time" was defined by combining the continuous variable "day" during the entire observation period (22 days for study 1; 31 days for study 2) and the dummy categorical variables to reflect "stage" (defined as baseline days 1-3, intervention days 4-17, and follow-up days 18-22 for study 1; and baseline days 1-10, intervention days 11-25, and follow-up days 26-31 for study 2) to allow different linear trends over different stages. The resulting 3 independent slopes represented the different changes of each group from baseline to the intervention stage, as well as from the intervention stage to the follow-up stage. The model also allowed different intercepts for each participant to control for possible variability in individuals' overall levels of steps. In this study, as participants

in all 3 groups could self-monitor their steps during the entire experimental period (baseline, intervention, and follow-up), the changes in the slopes from baseline to the intervention stage in the 2 social norm intervention groups will represent the extra effects of social norm information on walking (isolated from the effect of self-monitoring information in the entire experimental period).

Next, to more closely examine the intervention effect, we also created an LM model by only including the data from the intervention stage (14 days in study 1 and 15 days in study 2). The predictors of the model were group, day, and their interactions, with BMI as a covariate in study 1 and the participant as a random intercept term in the model. Then, by comparing the slopes of "day," we were able to compare the effects on walking performance between providing self-monitoring alone and providing social norm information and self-monitoring.

Considering that the ranking position may impact the effect of social norms, we also added ranking position as a new covariate in the PLM. The results remained the same with the ranking position as a covariate (see detailed information in Table S1 in Multimedia Appendix 1 and Table S1 in Multimedia Appendix 2).

We performed all data analyses using R 4.0.3 (R Project for Statistical Computing). We set the level of significance for all analyses to .05. A *P* value between .05 and .10 was considered marginally significant.

Results

Study 1

Participant Characteristics

Detailed information about participant characteristics is provided in Table 1. ANOVA showed that, for both males and females, control variables were not significantly different among groups, except for BMI (*P*=.03 and .03 for males and females, respectively). Consistent with our grouping criterion that the high-PA group would walk more, we observed a significantly higher baseline step count (male high-PA group: mean 11,775, SD 3970; female high-PA group: mean 10,860, SD 3879; *P*<.001), compared with the mean of the 3 low-PA groups. These results indicate that, except for BMI, the 3 low-PA groups were basically homogenous for both males and females. Thus, we added BMI as a covariate in all subsequent analyses for study 1.

Table 1. Participant demographic characteristics and baseline steps in study 1.

Variable	Male (N=54)				Female (N=63)			
	Self-monitoring (n=15)	PA[a]-consistent intervention (n=21)	PA-inconsistent intervention (n=18)	*P* value	Self-monitoring (n=22)	PA-consistent intervention (n=19)	PA-inconsistent intervention (n=22)	*P* value
Age (years), mean (SD)	25.87 (1.25)	25.57 (1.03)	25.72 (1.02)	.72	25.32 (0.89)	25.42 (0.96)	25.77 (1.02)	.27
Subjective health, mean (SD)	3.47 (0.92)	3.38 (0.86)	3.67 (0.69)	.55	3.32 (0.72)	3.32 (0.58)	3.05 (0.72)	.33
BMI, mean (SD)	22.54 (2.13)	22.86 (3.58)	20.51 (1.95)	.03	20.23 (2.29)	20.76 (1.23)	19.18 (1.83)	.03
Baseline steps, mean (SD)	8687 (3759)	7670 (3130)	8634 (3105)	.57	8356 (3717)	7527 (3407)	7623 (2376)	.66

[a]PA: physical activity.

Changes in Social Norms

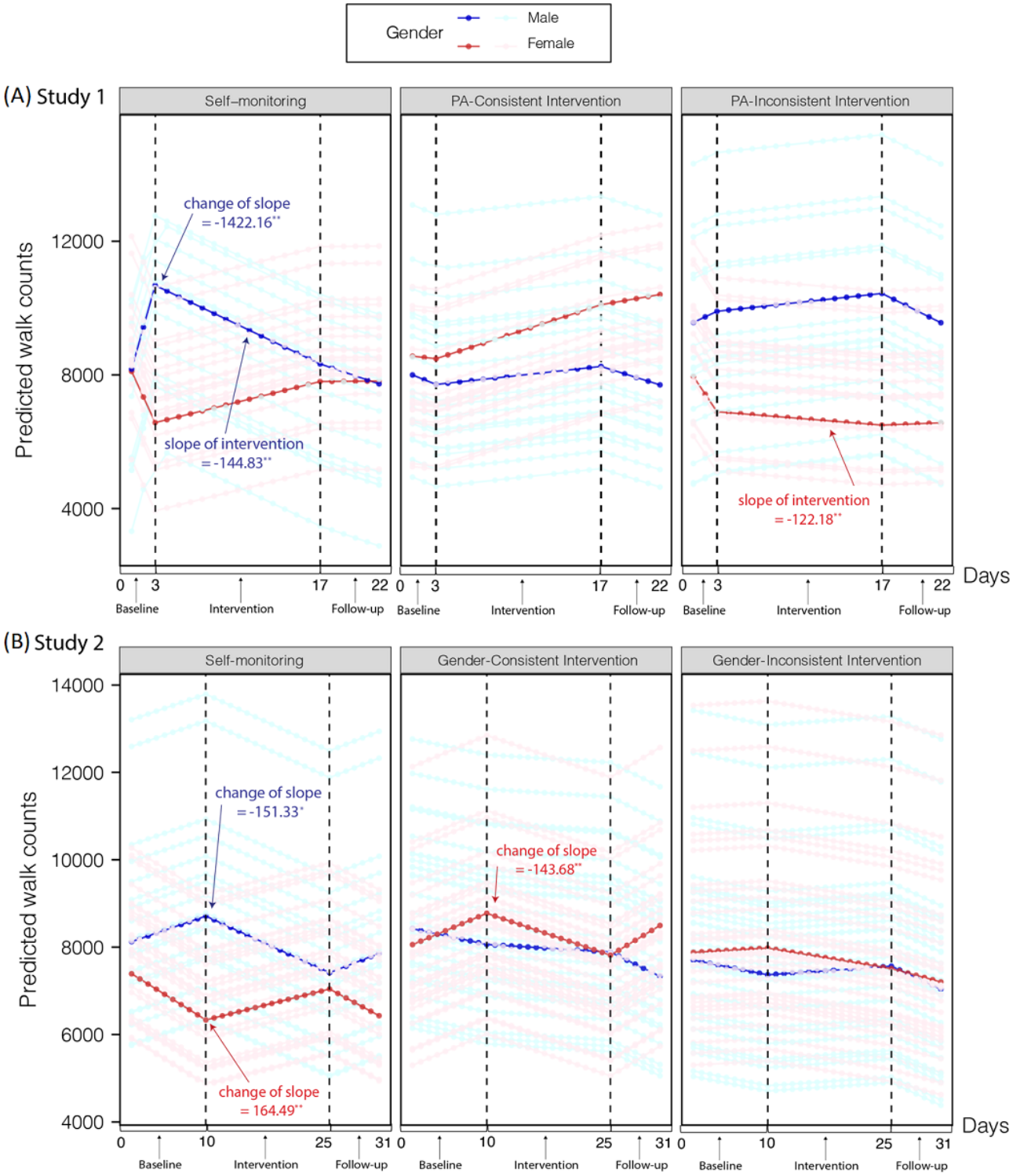
The LM model results (Table S1 in Multimedia Appendix 3) showed a significant main effect of stage on descriptive gender norms (male gender norm: $F_{2,222}$=28.70; *P*<.001; female gender norm: $F_{2,222}$=29.16; *P*<.001) and injunctive social norms ($F_{2,222}$=5.23; *P*=.01). In particular, during the experimental stages, the descriptive gender norm perceptions among all 3 groups increased progressively (*P*<.001). Participants' estimations about the percentage of male students walking over 6000 steps increased from a mean of 40.84% (SD 20.69%) to 49.80% (SD 18.62%) and then to 55.97% (SD 17.56%). Moreover, participants' estimations about the percentage of female students walking over 6000 steps increased from a mean of 31.83% (SD 19.32%) to 41.31% (SD 18.56%) and then to 47.21% (SD 17.85%). Among female students, injunctive social norm perceptions also increased significantly (*P*=.004) from baseline (mean 5.35, SD 0.92) to the follow-up stage (mean 5.66, SD 0.74). Surprisingly, however, we failed to observe a significant group difference in changes in social norms. This suggests that the self-monitoring intervention alone may be capable of leading people to perceive a higher level of social norms about walking.

Effect of the Social Norm Intervention on Step Count

As shown in Table 2 and Figure 3A, the PLM model showed separated patterns for genders with regard to step count. For males in the self-monitoring group, we detected a significantly negative change in slope over time from baseline to the intervention stage (*P*=.02), suggesting that they walked less after the 3-day baseline. For males within the 2 social norm intervention groups, walking did not change during the whole experiment. For females, however, no significant slopes or changes in slopes over time were found, suggesting that the

walking behavior of females did not change in any of the 3 groups during the whole experiment. BMI was also a significant predictor in females but not males. Specifically, a higher BMI predicted more steps taken by females (slope=328.13; $F_{1,59.23}$=6.21; P=.02), a pattern which was not repeated by males (slope=101.91; $F_{1,49.92}$=0.57; P=.45). These results suggest that higher BMI motivated females, but not males, to walk more steps.

Figure 3. Step counts (piece-wise linear mixed-effects model predicted) of the self-monitoring and 2 social norm intervention groups in study 1 (A) and study 2 (B). Each polyline represents a participant. One highlighted polyline is used for each group and gender (male, blue line; female, red line). All (marginally) significant changes of slope and slope of the intervention are labeled. *P<.10 but >.05, **P<.05.



These results partially supported H1 and H2, suggesting that the effect of social norm interventions is gender-specific. For males in the self-monitoring group, the promoting effect of self-monitoring on walking was short-lived. Although the social norm intervention did not robustly increase walking in males, the extra social norm information, regardless of its PA consistency, could at least prevent their walking steps from decreasing. For females, self-monitoring showed a stable but

nonsignificant effect on walking behavior, and we found no evidence that social norms worked better than self-monitoring.

Table 2. Results of the piece-wise linear mixed-effects model in study 1 across all experimental stages.

Gender and group	Intercept	Baseline		Intervention		Follow-up		Conditional R^2
		Slope	*P* value	Change in slope	*P* value	Change in slope	*P* value	
Male (N=54)								0.27
Self-monitoring (n=15)	8169.98	1254.20	.03	−1422.16	.02	48.66	.84	
PA[a]-consistent intervention (n=21)	7930.21	−142.53	.77	181.10	.71	−148.91	.46	
PA-inconsistent intervention (n=18)	9267.26	166.57	.75	−128.39	.82	−213.44	.32	
Female (N=63)								0.15
Self-monitoring (n=22)	8082.79	−766.83	.14	854.63	.12	−84.81	.68	
PA-consistent intervention (n=19)	8181.32	−37.13	.95	152.45	.79	−54.04	.81	
PA-inconsistent intervention (n=22)	8118.16	−525.53	.30	496.86	.35	42.54	.84	

[a]PA: physical activity.

Comparing the Effects of Self-Monitoring and Social Norms on Step Count

Results from the LM model at the intervention stage indicated (details in Table 3) a clear distinction between genders in terms of sensitivity to self-monitoring and social norm interventions. For males, the interaction between time and group was marginally significant ($F_{2,682.11}$=2.48; *P*=.09). In the self-monitoring group, we observed a significantly negative slope (*P*=.04), suggesting that steps taken by males in this group decreased during the intervention stage, while both the PA-consistent and PA-inconsistent interventions arrested this decreasing trend. For females, however, the interaction between time and group was significant ($F_{2,793.57}$=3.39; *P*=.03). A significantly negative slope (*P*=.03) for the PA-inconsistent intervention group also suggested that the PA-inconsistent intervention led to a decrease in steps walked during the intervention stage. No such effect for females in either the self-monitoring or PA-consistent intervention group was found. Consistent with findings from the PLM model, a higher BMI in females was a significant predictor of more steps walked ($F_{1,59.35}$=4.22; *P*=.04). Again, this pattern was not found in males ($F_{1,49.91}$=0.75; *P*=.39).

Table 3. Results of the intervention-focused linear mixed-effects model in study 1.

Gender and group	Intercept	Slope	SE	*P* value	Conditional R^2
Male (N=54)					0.31
Self-monitoring (n=15)	8448.02	−144.83	71.43	.04	
PA[a]-consistent intervention (n=21)	8068.35	37.04	60.92	.54	
PA-inconsistent intervention (n=18)	9520.74	48.70	65.76	.46	
Female (N=63)					0.21
Self-monitoring (n=22)	7800.66	81.14	57.49	.16	
PA-consistent intervention (n=19)	7973.10	32.73	61.70	.60	
PA-inconsistent intervention (n=22)	8141.32	−122.18	57.44	.03	

[a]PA: physical activity.

Taken together, these results partially supported H1 and H2, with several restrictions. For males in the self-monitoring group, self-monitoring only was not able to exert any long-term effect since steps decreased significantly, while for males in the 2 social norm intervention groups, steps were kept unchanged by the provision of social norm information. In contrast, for females, social norms did not work more effectively than self-monitoring. The hypothesized moderating effect of group identity (H3a) was supported, but it also interacted with gender. In particular, for females, PA-inconsistent social norms had a negative effect on walking behavior.

In summary, these results preliminarily confirmed H2 and H3, revealing a gender-specific moderating effect of group identity on social norms. Given the finding that gender did play an

essential role in moderating the effect of the social norm intervention and considering the reliability of our results, we conducted study 2 to replicate this study. Moreover, we modulated gender consistency to further examine the effect of group identity.

Study 2

Participant Characteristics

Detailed information of participant characteristics can be found in Table 4. ANOVA showed that, for both males and females, these baseline characteristics were not significantly different among groups (Table 4), indicating a fair homogeneity for the 3 experimental groups among both males and females.

Table 4. Participant demographic characteristics and baseline steps in study 2.

Variable	Male (N=88)				Female (N=92)			
	Self-monitoring (n=23)	Gender-consistent intervention (n=33)	Gender-inconsistent intervention (n=32)	*P* value	Self-monitoring (n=26)	Gender-consistent intervention (n=34)	Gender-inconsistent intervention (n=32)	*P* value
Age (years), mean (SD)	22.57 (1.20)	23.06 (1.50)	22.88 (1.29)	.41	22.58 (0.95)	22.24 (1.02)	22.28 (0.63)	.29
Subjective health, mean (SD)	3.30 (0.82)	3.42 (0.61)	3.41 (0.50)	.77	3.35 (0.56)	3.29 (0.72)	3.44 (0.56)	.65
BMI, mean (SD)	22.40 (3.56)	21.71 (2.52)	21.41 (3.72)	.54	19.89 (2.59)	19.42 (1.16)	20.54 (2.52)	.14
Baseline steps, mean (SD)	8737 (2279)	8750 (2211)	8012 (2647)	.39	7358 (1518)	8099 (1732)	8307 (2788)	.22

Changes in Social Norms

Similar to study 1 (Table S1 in Multimedia Appendix 4), for the 2 types of descriptive gender norms, LM model results showed a significant main effect of the stage (male gender norm: $F_{2,348}$=13.80; P<.001; female gender norm: $F_{2,348}$=22.39; P<.001). Post-hoc analysis showed that both descriptive male and female norms increased (P<.001) as the experiment progressed (from baseline to intervention to follow-up). However, in terms of injunctive social norms, LM model results showed no significant main or interaction effect. These findings indicate that injunctive norms in participants across all 3 groups did not change significantly with the experiment process.

In contrast to study 1, we found that females perceived higher descriptive gender norms than males. When asked "What is your estimation of the percentage of male/female students in the same university walking over 6000 steps?" (descriptive male/female gender norms), female participants estimated that 52.53% (SD 20.68%) of male students and 40.13% (SD 19.76%) of female students would walk over 6000 steps, while male participants estimated that 42.00% (SD 19.98%) of male students and 33.99% (SD 19.06%) of female students would walk over 6000 steps. In addition, there was a significant interaction between group and gender (P=.01) for the descriptive male norm. Post-hoc analysis showed that there was no significant group difference for female participants (adjusted P>.10). However, for male participants, the gender-consistent group had significantly higher norms than both the self-monitoring (adjusted P=.01) and PA-consistent groups (adjusted P=.09). In detail, males in the gender-consistent group estimated that 48.29% (SD 19.91%) of male students walked over 6000 steps, while male participants in the self-monitoring and gender-inconsistent groups estimated that 35.54% (SD 19.51%) and 40.16% (SD 18.67%) of male students walked over 6000 steps, respectively. In contrast, their estimations for female students were not significantly different.

Again, these results indicated that descriptive gender norms kept increasing for both genders as the experiment progressed from baseline to intervention to follow-up. We found that the gender-consistent manipulation was effective in influencing the gender norm perceptions of males but not females.

Effect of the Social Norm Intervention on Walking Behavior

Results of the PLM model in Table 5 and Figure 3B replicated the finding in study 1 of a different effect of self-monitoring and social norms for each gender. For males, a self-monitoring–only intervention led them to walk less after baseline, while additional social norm information may help keep their walk steps unchanged. Only males in the self-monitoring group walked less after baseline as indicated by a marginally significant negative change of slope over time from baseline to the intervention stage (P=.08). On the other hand, males in the 2 social norm intervention groups did not show any change in walking behavior during the entire experiment.

However, in contrast to study 1, we found a backfire effect of gender-consistent social norms among females. Females in the self-monitoring group walked more after baseline, and the slope changed in a positive direction over time from baseline to the intervention stage (P=.03). However, females in the gender-consistent group walked less compared to baseline (ie, the backfire effect), and the slope changed in a negative direction over time from baseline to the intervention stage (P=.03). These results again suggested that social norms might help males to maintain long-term walking behaviors, while they do not provide the same benefit and could even exert a negative effect

(gender-consistent social norms) on the walking behavior of females.

Table 5. Results of the piece-wise linear mixed-effects model in study 2 across all experimental stages.

Gender and group	Intercept	Baseline		Intervention		Follow-up		Conditional R^2
		Slope	*P* value	Change in slope	*P* value	Change in slope	*P* value	
Male (N=88)								0.25
Self-monitoring (n=23)	8159.99	64.89	.30	−151.33	.08	160.51	.20	
Gender-consistent intervention (n=33)	8518.26	−39.97	.45	28.29	.69	−81.14	.44	
Gender-inconsistent intervention (n=32)	7839.48	−37.16	.49	50.25	.49	−100.37	.35	
Female (N=92)								0.22
Self-monitoring (n=26)	7303.92	−117.33	.04	164.49	.03	−148.91	.19	
Gender-consistent intervention (n=34)	8014.41	79.70	.11	−143.68	.03	177.38	.07	
Gender-inconsistent intervention (n=32)	7970.47	10.94	.83	−42.36	.54	−19.19	.85	

Comparing the Effects of Self-Monitoring and Social Norms on Walking Behavior

We did not find any significant slopes in the intervention-focused model. Detailed results from the LM model over the intervention period are shown in Table S1 and S2 in Multimedia Appendix 2. These results partially supported H1 and H2, and essentially replicated results from study 1, suggesting that the effect of social norm interventions on walking is only weak among males but not females. For males, walking behaviors that were driven by self-monitoring only began to decrease slightly after 10 days, while for those receiving social norm information, regardless of whether this information was gender-specific (ie, in the gender-consistent group) or not (ie, in the gender-inconsistent group), the walking behaviors remained unchanged. However, contradicting H3b, for females, gender-consistent social norms produced a backfire effect on the promotion of walking, suggesting that without these social norms, self-monitoring only may have led to females walking more.

Discussion

Principal Findings

Through 2 longitudinal tracking studies in a sample of Chinese college students, we made several findings. First, gender moderated the effect of the social norm intervention on walking. For males, the effect of the self-monitoring–only intervention was short-lived, while the addition of social norm information, regardless of its PA consistency or gender consistency, was able to keep the walking behavior unchanged. For females, we found no evidence to show that social norms performed better than self-monitoring. Second, an additional higher level of group identity in females did not consistently guarantee an improved effect of the social norm intervention on PA. With PA-inconsistent social norms (study 1) and gender-consistent social norms (study 2), the results suggested even a backfire effect on walking steps among females.

Gender-Specific Effectiveness of Social Norm Interventions for Walking

Most previous intervention studies considered self-monitoring and social norms as 2 independent and effective strategies for promoting PA [7,8,12]. Little attention was paid to how these 2 strategies might compare in terms of effectiveness. In our studies, we combined these 2 strategies (self-monitoring–only intervention vs social norms and self-monitoring intervention) in order to directly compare their effects. Consistent with previous work [10,12,26], both studies consistently revealed that the effects of self-monitoring on PA (especially among males) were short-term (at least no longer than 3 or 10 days). We found that, among males, the addition of sharing social norms could outperform the effects of self-monitoring only. These results were partially consistent with those of Rote et al [9]. They reported that young women who shared PA information with Facebook group members were able to increase their PA more than those who only used a self-monitoring strategy. These results differed from our results in terms of gender.

As a meta-analysis [23] has previously suggested, our results revealed a notable separation in the patterns of PA outcomes according to gender. Specifically, for males, with self-monitoring as a baseline condition, we found that the effect of the intervention with social norm information can keep walking unchanged among males. However, for females, we found no evidence that social norms performed better than self-monitoring. A reasonable explanation for gender differences in the effect of a social norm intervention is that a relatively weak self-monitoring effect among males may leave more potential for a social norm intervention to produce a protective effect. Previous studies suggested that male students may be less self-disciplined [64] and less likely to use self-regulation strategies [65] than females. As a result, both of our studies

showed a marginally significant decrease in walking in males after the baseline stage (in study 1, the change of slope over time was −1422.16; *P*=.02; in study 2, the change of slope over time was −151.33; *P*=.08), revealing a short-lived effect of self-monitoring among males when no social norm information was provided. Thus, as the effect of self-monitoring began to fade over time [66], additional social norm information was able to encourage males to perform social comparisons between themselves and their peers, thus motivating them to pursue a better rank position by walking more. For females, however, the self-monitoring intervention tended to benefit them more significantly, as they represent a more self-disciplined group [64]. A previous meta-analysis also concluded that the effect size of wearing a pedometer is much greater for females (95% CI 0.64-0.97) than males (95% CI −0.18 to 0.79) [66]. Therefore, a simple wearable pedometer may already lead highly motivated females to walk more, thus making it more difficult for other interventions to produce an additional enhancement or further increase the PA of females. This may be the reason why social norm information or supplemental PA intervention strategies [67], other than self-monitoring, work better for males than females.

An alternative explanation may be related to a counterproductive cultural stereotype, prevalent in Chinese culture, that females should be less active and petite than males [68]. From this perspective, constant PA may be deemed as an obstacle to adhering to such a stereotype regarding femininity, which may consequently discourage some female college students from engaging in PA. This explanation is supported by our results that the descriptive female gender norms were consistently lower than the male gender norms in both studies. Taking study 2 as an example, participants estimated that only 37.13% (SD 19.65%) of female students would walk over 6000 steps; however, for male students, the estimation was 47.38% (SD 20.99%). Still, these stereotypes of walking gender difference may only be effective as a self-serving bias for males. In study 2, the actual within-group rankings of females were even better than those of males; 53% of the highest quarter of the step ranking involved females in the gender-inconsistent intervention group. As a result, a relatively high level of individual PA may be perceived as a quality outside the “normal” confines of femininity, thus deviating from the traditional female stereotype, or, even worse, it could be considered as having an atypical female behavior. Previous studies have shown that a high level of PA may expose females to feelings of social pressure [69] and being stigmatized [70]. Thus, even if a female could have a similar step ranking to a male, the psychological meaning of ranking (the social norm information) might be different because, among females, having a high PA level might be perceived as an outside-gender group behavior rather than a within-gender group social norm. However, males may be motivated purely by the desire to pursue a better rank position, since walking more or having high PA is consistent with their own gender stereotype and set of social expectations. Thus, ranking as a motivational device was only effective for males, and in light of cultural expectations regarding femininity, no positive effect of social norm information was found among females. Future studies may benefit from comparing the effects of self-monitoring and social norm interventions on PA within different cultural backgrounds, or perhaps examining gender stereotypes unique to the cultures of participants’ countries.

It should be pointed out that, based on the current evidence, providing social norm information might have no more than a protective effect. Since neither study revealed a significantly positive change in slope in any social norm intervention group during the intervention stage, the provision of social norm information did not produce any additional enhancement in PA. Among males, however, both studies suggested that self-monitoring on its own may lead to a reduction in steps walked following the baseline stage. The addition of social norm information was, at least, shown to hold the level of PA steady in males, which can be considered as a protective effect.

A Higher Group Identity Does Not Equal A Better Effect of Social Norm Information

In both studies, we found that group identity moderated the effect of social norm information on walking behavior. According to predictions of the TNSB, social norm information shared between members with a high group identity should make its effect more powerful [42]. However, we found that this type of effect of group identity was not reflected among males, while, among females, a higher group identity appeared to be a double-edged sword in terms of its effect on PA.

In particular, in study 1, we found that providing social norms with lower group identity (PA-inconsistent social norms) led to a reduced effect on walking in females, a finding which is consistent with the TNSB. This finding among females who were part of the PA-inconsistent group engaging in a low level of PA may be driven by 2 possible factors. First, we have to note the large disparity in performance between high-PA participants and low-PA participants in our study. For high-PA participants, the mean count of steps walked during the intervention stage was 11,997 (SD 3165), but for low-PA participants, the mean count of steps walked was 7868 (SD 2341). The later participants may have been likely to perceive themselves as outgroup members, thus weakening the effect of any social norm intervention. Second, these members may also have perceived a greater sense of upward social comparison, that is, comparison with peers who are relatively better off [71]. Thereafter, there may have been a subsequent loss of motivation to achieve a better ranking, leading to a reduction in walking. This effect is consistent with previous studies in which more upward comparison on social media platforms was associated with a decrease in self-esteem or well-being [72], as well as more depression or shame [73]. It has been suggested that interventions that lead to inappropriate types of social comparison, either upward or downward, may indeed be counter effective for one’s PA. Further studies should take the type of social comparison into account when considering the application of other potential factors (eg, age consistency [46]) to increase (or decrease) group identity theoretically.

In study 2, however, social norm information from a gender-consistent source, which should theoretically lead to a higher group identity, surprisingly led to a decrease in PA among females compared with PA in the self-monitoring control group. Although there are few interventions using gender-consistent social norms, our results are consistent with the finding in

another health intervention study (providing gender-consistent social norm information is ineffective in reducing alcohol intake) [47]. One possible explanation might be that for females, sharing gender-consistent social norm information, that is, sharing gender-specific information, may provoke increased feelings associated with femininity and thus bring about concerns regarding their own social gender role. An alternative explanation could be that the separation of genders in ranking may unintentionally sharpen some undesirable social norms for females (eg, few females walk a lot), thus negatively impacting the intervention. For future interventions, we would argue that, at least in the context of promoting PA among females, tailoring social norm interventions to be gender consistent may not be advisable, since an unexpected side effect may conceal or even reverse the desired effect of the interventions.

Implications

Based on WeChat groups and step ranking information as an ecological intervention, our study can provide new empirical evidence with high ecological validity regarding the effect of social norm information in promoting PA within the Chinese context. Our study contributes to the field of mHealth and PA interventions in 3 ways. First, by using a self-monitoring control group, our study avoided the high statistical error rate typically brought about by a mixed-strategy intervention. Second, our findings highlighted salient gender differences associated with the effects of social norm information and self-monitoring on PA. These findings suggest the necessity to take gender into account when designing mHealth interventions for Chinese users. Third, we advise caution in using group identity to promote PA. This is especially true for gender-specific manipulations, since the increase in identification with females may, in the context of promoting walking or general PA, conflict with the target behavior.

Given the critical role of personalization in mHealth interventions, our results provide some useful insights into PA-targeted mHealth intervention projects in China. First, a gender-specialized intervention is essential to the goal of nudging walking behavior in China. For males, compared with a more traditional passive presentation of walking data, proactively pushing PA-related social norm information to users may be a more effective method for pedometer apps and other self-monitoring interventions. For example, WeRun or other mHealth apps could push a daily reminder containing the walking data of the friends of male users in order to maintain the initial beneficial effect of self-monitoring. However, among females, since social norm information did not have a beneficial effect on walking behavior, rather than adding extra social norm manipulations, maximizing the effect of self-monitoring may be a better approach. Female-targeted interventions should make full use of self-monitoring strategies. For example, they could push a daily personal step count to female users and track the overall walking step count record for each week. Second, nonessential gender-specific social norm information should be used with caution in PA interventions, since heightened gender identity may conflict with the target behavior, thus leading to an undesired outcome. Therefore, further mHealth interventions or smart devices should tailor their plans according to the gender of users; the provision of social norm information may be “the icing on the cake” for males but would be superfluous for females.

Limitations

Several limitations of this research should be noted. All of our participants were Chinese college students, and this may limit the external validity and generalizability of our results. Although the main results were replicated across 2 studies to ensure their reliability, more empirical evidence is needed before generalizing our findings to other samples (eg, older adults) or other cultural backgrounds (eg, those with relatively weak gender stereotypes or gender gaps). Given that college students have a relatively high PA level (the average step count recorded by WeRun was 6932 steps/day in 2019 [74], while participants reached 7900 step/day in both our studies), further studies may investigate lower PA groups, such as office staff and retirees, to replicate our results.

In addition, previous studies have suggested that the effect of self-monitoring on PA can be attributed to the Hawthorne effect [26]. Although we aimed to compare the pure effect of self-monitoring and a social norm intervention, the nature of our study design means that we cannot rule out this possibility, that is, any effects observed may have been related to participants’ awareness that their walking activity was being monitored.

Finally, we did not provide sufficient empirical evidence to attribute our finding that females reduced their PA in the gender-consistent group to the presence of a backfire effect of high group identity on social norms. Other potential variables, such as trait competitiveness, may also impact the effect of social norm information. Further studies may benefit from controlling these variables to directly address this unexplained result.

Conclusion

Gender and group identity are 2 moderators of the effect of social norm interventions on walking. In female Chinese college students, a higher sense of group identity does not guarantee a better effect of social norm information on PA, that is, PA-inconsistent or gender-consistent social norm information reduces walking in females. However, in male Chinese college students, social norm information and group identity do not show such an effect. Specifically, while walking may decrease with self-monitoring only, the inclusion of social norm information may help to keep the level of walking steady in males.

Acknowledgments

We would like to thank J Murray and S Lippke for their constructive comments during peer review. We also acknowledge David O’Connor for his language editing. We also thank Wang Xiao-Ming, Chen Jun-Fang, Yang Shu-Wen, and Sui Xiao-Yang for

their help during data collection. This research was supported by Beijing Natural Science Foundation (BNSF, 9172019), the National Natural Science Foundation of China (NSFC, 71471171), CAS Engineering Laboratory for Psychological Service (KFJ-PTXM-29), the Scientific Foundation of the Institute of Psychology, Chinese Academy of Sciences (Y9CX303008), and Major Projects of the National Social Science Foundation of China (19ZDA358).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional analysis of the effect of self-monitoring and social norm information on walking behavior (study 1).
[DOCX File , 39 KB - jmir_v23i12e29167_app1.docx]

Multimedia Appendix 2
Additional analysis of the effect of self-monitoring and social norm information on walking behavior (study 2).
[DOCX File , 40 KB - jmir_v23i12e29167_app2.docx]

Multimedia Appendix 3
Descriptive information of social norms (study 1).
[DOCX File , 39 KB - jmir_v23i12e29167_app3.docx]

Multimedia Appendix 4
Descriptive information of social norms (study 2).
[DOCX File , 39 KB - jmir_v23i12e29167_app4.docx]

References

1. Melzer D, Gardener E, Guralnik JM. Mobility disability in the middle-aged: cross-sectional associations in the English Longitudinal Study of Ageing. Age Ageing 2005 Nov;34(6):594-602. [doi: 10.1093/ageing/afi188] [Medline: 16267185]
2. Blair SN. Physical Fitness and All-Cause Mortality. JAMA 1989 Nov 03;262(17):2395. [doi: 10.1001/jama.1989.03430170057028]
3. Pate RR. Physical Activity and Public Health. JAMA 1995 Feb 01;273(5):402. [doi: 10.1001/jama.1995.03520290054029]
4. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
5. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
6. Noland MP. The effects of self-monitoring and reinforcement on exercise adherence. Res Q Exerc Sport 1989 Sep;60(3):216-224. [doi: 10.1080/02701367.1989.10607443] [Medline: 2489846]
7. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
8. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. Ann Fam Med 2008 Jan 01;6(1):69-77 [FREE Full text] [doi: 10.1370/afm.761] [Medline: 18195317]
9. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The Efficacy of a Walking Intervention Using Social Media to Increase Physical Activity: A Randomized Trial. J Phys Act Health 2015 Jun 16;12 Suppl 1:S18-S25. [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]
10. 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]
11. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013 Apr;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
12. Priebe CS, Spink KS. Using messages promoting descriptive norms to increase physical activity. Health Commun 2012 Apr;27(3):284-291. [doi: 10.1080/10410236.2011.585448] [Medline: 21899404]
13. Priebe CS, Spink KS. When in Rome: Descriptive norms and physical activity. Psychology of Sport and Exercise 2011 Mar;12(2):93-98. [doi: 10.1016/j.psychsport.2010.09.001]

14. Clapp JD, Lange JE, Russell C, Shillington A, Voas RB. A failed norms social marketing campaign. J Stud Alcohol 2003 May;64(3):409-414. [doi: 10.15288/jsa.2003.64.409] [Medline: 12817831]
15. Moreira MT, Smith LA, Foxcroft D. Social norms interventions to reduce alcohol misuse in university or college students. Cochrane Database Syst Rev 2009 Jul 08(3):CD006748. [doi: 10.1002/14651858.CD006748.pub2] [Medline: 19588402]
16. Kim J, Dunn E, Rellinger K, Robertson-Wilson J, Eys M. Social norms and physical activity in American and Canadian contexts: a scoping review. International Review of Sport and Exercise Psychology 2017 Jul 27;12(1):26-48. [doi: 10.1080/1750984x.2017.1354229]
17. Kanfer FH. Self-monitoring: Methodological limitations and clinical applications. Journal of Consulting and Clinical Psychology 1970;35(2):148-152. [doi: 10.1037/h0029874]
18. Kanfer FH, Gaelick-Buys L. Self-management methods. In: Kanfer FH, Goldstein AP, editors. Helping people change: A textbook of methods. Oxford, United Kingdom: Pergamon Press; 1991:305-360.
19. Frederiksen LW, Epstein LH, Kosevsky BP. Reliability and Controlling Effects of Three Procedures for Self-Monitoring Smoking. Psychol Rec 2017 Jun 1;25(2):255-263. [doi: 10.1007/bf03394311]
20. Turnin MG, Beddok RH, Clottes JP, Martini PF, Abadie RG, Buisson J, et al. Telematic expert system Diabeto. New tool for diet self-monitoring for diabetic patients. Diabetes Care 1992 Feb 01;15(2):204-212. [doi: 10.2337/diacare.15.2.204] [Medline: 1547677]
21. Swartz AM, Strath SJ, Bassett DR, Moore J, Redwine BA, Groër M, et al. Increasing daily walking improves glucose tolerance in overweight women. Preventive Medicine 2003 Oct;37(4):356-362. [doi: 10.1016/s0091-7435(03)00144-0]
22. Kanejima Y, Kitamura M, Izawa KP. Self-monitoring to increase physical activity in patients with cardiovascular disease: a systematic review and meta-analysis. Aging Clin Exp Res 2019 Feb 30;31(2):163-173. [doi: 10.1007/s40520-018-0960-7] [Medline: 29714027]
23. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Sep 15;12(3):211-230. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]
24. Baker RC, Kirschenbaum DS. Self-monitoring may be necessary for successful weight control. Behavior Therapy 1993;24(3):377-394. [doi: 10.1016/S0005-7894(05)80212-6]
25. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
26. Butler L, Dwyer D. Pedometers may not provide a positive effect on walking activity. Health Promot J Aust 2004 Aug 01;15(2):134-136. [doi: 10.1071/he04134]
27. Vašíčková J, Groffik D, Frömel K, Chmelík F, Wasowicz W. Determining gender differences in adolescent physical activity levels using IPAQ long form and pedometers. Ann Agric Environ Med 2013;20(4):749-755 [FREE Full text] [Medline: 24364448]
28. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WMP, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
29. Cialdini RB, Goldstein NJ. Social influence: compliance and conformity. Annu Rev Psychol 2004 Feb;55:591-621. [doi: 10.1146/annurev.psych.55.090902.142015] [Medline: 14744228]
30. Cialdini RB, Reno RR, Kallgren CA. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology 1990;58(6):1015-1026. [doi: 10.1037/0022-3514.58.6.1015]
31. Mollen S, Rimal RN, Ruiter RA, Jang SA, Kok G. Intervening or interfering? The influence of injunctive and descriptive norms on intervention behaviours in alcohol consumption contexts. Psychol Health 2013 May;28(5):561-578. [doi: 10.1080/08870446.2012.752827] [Medline: 23289490]
32. Ranby KW, Aiken LS. Incorporating husband influences into a model of physical activity among older women. Br J Health Psychol 2016 Sep 09;21(3):677-693. [doi: 10.1111/bjhp.12195] [Medline: 27159040]
33. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. Health Psychology Review 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
34. Sheeran P, Harris PR, Epton T. Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. Psychol Bull 2014 Mar;140(2):511-543. [doi: 10.1037/a0033065] [Medline: 23731175]
35. Hagger MS, Chatzisarantis NLD, Biddle SJH. A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: Predictive validity and the contribution of additional variables. Journal of Sport & Exercise Psychology 2002;24(1):3-32. [doi: 10.1123/jsep.24.1.3]
36. Rhodes RE, Nigg CR. Advancing Physical Activity Theory. Exercise and Sport Sciences Reviews 2011;39(3):113-119. [doi: 10.1097/jes.0b013e31821b94c8]
37. Stok FM, de Vet E, de Ridder DT, de Wit JB. The potential of peer social norms to shape food intake in adolescents and young adults: a systematic review of effects and moderators. Health Psychol Rev 2016 Sep 22;10(3):326-340. [doi: 10.1080/17437199.2016.1155161] [Medline: 26878931]

38. Miller DT, Prentice DA. Changing Norms to Change Behavior. Annu Rev Psychol 2016 Jan 04;67(1):339-361. [doi: 10.1146/annurev-psych-010814-015013] [Medline: 26253542]
39. Chapman GB, Colby H, Convery K, Coups EJ. Goals and Social Comparisons Promote Walking Behavior. Med Decis Making 2016 May;36(4):472-478. [doi: 10.1177/0272989X15592156] [Medline: 26139447]
40. Jang SA, Rimal RN, Cho N. Normative influences and alcohol consumption: the role of drinking refusal self-efficacy. Health Commun 2013;28(5):443-451. [doi: 10.1080/10410236.2012.691455] [Medline: 22809467]
41. Lapinski MK, Anderson J, Shugart A, Todd E. Social influence in child care centers: a test of the theory of normative social behavior. Health Commun 2014 May 17;29(3):219-232. [doi: 10.1080/10410236.2012.738322] [Medline: 23682754]
42. Rimal RN, Real K. How Behaviors are Influenced by Perceived Norms. Communication Research 2016 Jun 29;32(3):389-414. [doi: 10.1177/0093650205275385]
43. Godfrey AS, Niekitta Z. The Influence of Groupthink on Culture and Conflict in Twitter. In: Reconceptualizing New Media and Intercultural Communication in a Networked Society. Hershey, PA: IGI Global; 2018.
44. Neighbors C, LaBrie JW, Hummer JF, Lewis MA, Lee CM, Desai S, et al. Group identification as a moderator of the relationship between perceived social norms and alcohol consumption. Psychol Addict Behav 2010 Sep;24(3):522-528 [FREE Full text] [doi: 10.1037/a0019944] [Medline: 20853938]
45. Lewis MA, Neighbors C. Optimizing personalized normative feedback: the use of gender-specific referents. J Stud Alcohol Drugs 2007 Mar;68(2):228-237 [FREE Full text] [doi: 10.15288/jsad.2007.68.228] [Medline: 17286341]
46. Lewis MA, Neighbors C. Gender-Specific Misperceptions of College Student Drinking Norms. Psychology of Addictive Behaviors 2004;18(4):334-339. [doi: 10.1037/0893-164x.18.4.334]
47. Lojewski R, Rotunda RJ, Arruda JE. Personalized Normative Feedback to Reduce Drinking among College Students: A Social Norms Intervention Examining Gender-based Versus Standard Feedback. Journal of Alcohol and Drug Education 2010;54(3):19-40 [FREE Full text]
48. Conner M, Smith N, McMillan B. Examining normative pressure in the theory of planned behaviour: Impact of gender and passengers on intentions to break the speed limit. Curr Psychol 2003 Sep;22(3):252-263. [doi: 10.1007/s12144-003-1020-8]
49. Gerber JP. Social Comparison Theory. In: Zeigler-Hill V, Shackelford TK, editors. Encyclopedia of Personality and Individual Differences. Cham, Switzerland: Springer; 2018.
50. Pollard TM, Wagnild JM. Gender differences in walking (for leisure, transport and in total) across adult life: a systematic review. BMC Public Health 2017 Apr 20;17(1):341 [FREE Full text] [doi: 10.1186/s12889-017-4253-4] [Medline: 28427376]
51. Hyun-Woo J, Joo A, Hak-Seon K. Eating behavior and physical activity among college students: A descriptive approach to the gender difference. Journal of Culinary Research 2014;20(5):16-21 [FREE Full text]
52. Magoc D, Tomaka J, Shamaley AG, Bridges A. Gender Differences in Physical Activity and Related Beliefs Among Hispanic College Students. Hispanic Journal of Behavioral Sciences 2016 Mar 14;38(2):279-290. [doi: 10.1177/0739986316637355]
53. McArthur L, Raedeke TD. Race and sex differences in college student physical activity correlates. Am J Health Behav 2009;33(1):80-90. [doi: 10.5993/ajhb.33.1.8] [Medline: 18844523]
54. The Global Gender Gap Index 2020 Rankings. World Economic Forum. URL: http://reports.weforum.org/global-gender-gap-report-2020/the-global-gender-gap-index-2020-rankings/ [accessed 2021-11-29]
55. Li J, Luo X. Gap vs. Effort: Comparative Study of Sports Attitude of College Students from China and US. TOCSJ 2015 Oct 28;9(1):2587-2591. [doi: 10.2174/1874110x01509012587]
56. Shane T, Wang N. Gender differences in physical training behavior of college students. Journal of Shenyang Sports University 2007;26(1):103-106.
57. China Mobile Medical User Scale and Forecast. iimedia. URL: https://data.iimedia.cn/page-category.jsp?nodeid=13210376 [accessed 2021-11-29]
58. Fong SS, Ng SS, Cheng YT, Zhang J, Chung LM, Chow GC, et al. Comparison between smartphone pedometer applications and traditional pedometers for improving physical activity and body mass index in community-dwelling older adults. J Phys Ther Sci 2016 May;28(5):1651-1656 [FREE Full text] [doi: 10.1589/jpts.28.1651] [Medline: 27313391]
59. Montag C, Becker B, Gan C. The Multipurpose Application WeChat: A Review on Recent Research. Front Psychol 2018;9:2247 [FREE Full text] [doi: 10.3389/fpsyg.2018.02247] [Medline: 30618894]
60. WeChat impact report 2018: all the latest WeChat data. WalktheChat. 2018. URL: https://walkthechat.com/wechat-impact-report-2016/ [accessed 2021-11-29]
61. Primack BA, Switzer GE, Dalton MA. Improving measurement of normative beliefs involving smoking among adolescents. Arch Pediatr Adolesc Med 2007 May 01;161(5):434-439 [FREE Full text] [doi: 10.1001/archpedi.161.5.434] [Medline: 17485617]
62. Corn, BW, Marcus, SM, Topham A, Hauck, W, Curran WJ. Will primary central nervous system lymphoma be the most frequent brain tumor diagnosed in the year 2000? Cancer 1997 Jun 15;79(12):2409-2413. [doi: 10.1002/(sici)1097-0142(19970615)79:12<2409::aid-cncr17>3.0.co;2-v]
63. Lorch RF, Myers JL. Regression analyses of repeated measures data in cognitive research. Journal of Experimental Psychology: Learning, Memory, and Cognition 1990;16(1):149-157. [doi: 10.1037/0278-7393.16.1.149]

64. Duckworth AL, Seligman MEP. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of Educational Psychology 2006;98(1):198-208. [doi: 10.1037/0022-0663.98.1.198]
65. Bembenutty H. Academic delay of gratification, self-regulation of learning, gender differences, and expectancy-value. Personality and Individual Differences 2009 Feb;46(3):347-352. [doi: 10.1016/j.paid.2008.10.028]
66. Kang M, Marshall SJ, Barreira TV, Lee J. Effect of Pedometer-Based Physical Activity Interventions: A Meta-Analysis. Res Quart Exerc Sport 2009 Sep 01;80(3):648-655. [doi: 10.5641/027013609x13088500160000]
67. Liau AK, Neihart M, Teo CT, Goh LS, Chew P. A Quasi-Experimental Study of a Fitbit-Based Self-Regulation Intervention to Improve Physical Activity, Well-Being, and Mental Health. Cyberpsychol Behav Soc Netw 2018 Nov;21(11):727-734. [doi: 10.1089/cyber.2016.0502] [Medline: 30421998]
68. Yang J, Ni XH, Tang L, Duan AM. The impact of sports stereotypes on the sports participation of undergraduate female students. Sport Science and Technology 2016;37(6):39-40. [doi: 10.14038/j.cnki.tykj.2016.06.095]
69. Spencer RA, Rehman L, Kirk SF. Understanding gender norms, nutrition, and physical activity in adolescent girls: a scoping review. Int J Behav Nutr Phys Act 2015 Jan 24;12(1):6 [FREE Full text] [doi: 10.1186/s12966-015-0166-8] [Medline: 25616739]
70. Yang YH, Wu GL. Origins of gender-based stereotypes and "stigmatization" of female and masculine athletes. Collection of Women's Studies 2014;1:60-66.
71. Collins RL. For better or worse: The impact of upward social comparison on self-evaluations. Psychological Bulletin 1996;119(1):51-69. [doi: 10.1037/0033-2909.119.1.51]
72. Schmuck D, Karsay K, Matthes J, Stevic A. “Looking Up and Feeling Down”. The influence of mobile social networking site use on upward social comparison, self-esteem, and well-being of adult smartphone users. Telematics and Informatics 2019 Sep;42:101240 [FREE Full text] [doi: 10.1016/j.tele.2019.101240]
73. Tosun LP, Kaşdarma E. Passive Facebook Use and Depression. Journal of Media Psychology 2020 Oct 16;32(4):165-175. [doi: 10.1027/1864-1105/a000269]
74. 2019 WeChat Data Report. weixin. URL: https://mp.weixin.qq.com/s/vmhoiRzpBs7-JK_x2a7gZw [accessed 2021-05-12]

Abbreviations

ANOVA: analysis of variance
LM model: linear mixed-effects model
mHealth: mobile health
PA: physical activity
PLM model: piece-wise linear mixed-effects model
TNSB: theory of normative social behavior

Edited by R Kukafka; submitted 05.06.21; peer-reviewed by J Murray, S Lippke; comments to author 30.07.21; revised version received 23.10.21; accepted 29.10.21; published 07.12.21.

Please cite as:
Xu Y, Yue LZ, Wang W, Wu XJ, Liang ZY
Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study
J Med Internet Res 2021;23(12):e29167
URL: https://www.jmir.org/2021/12/e29167
doi:10.2196/29167
PMID:34878992

Original Paper

Condomless Anal Sex Associated With Heterogeneous Profiles Of HIV Pre-Exposure Prophylaxis Use and Sexual Activities Among Men Who Have Sex With Men: A Latent Class Analysis Using Sex Diary Data on a Mobile App

Yi-Fang Yu[1], MSc; Huei-Jiuan Wu[2], MSc; Stephane Wen-Wei Ku[3], MD; Po-Hsien Huang[4], PhD; Chia-Wen Li[5], MSc, MD; Poyao Huang[6], PhD; Carol Strong[1], PhD

[1]Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[2]The Kirby Institute, University of New South Wales, Sydney, Australia

[3]Division of Infectious Diseases, Department of Medicine, Taipei City Hospital Renai Branch, Taipei, Taiwan

[4]Department of Psychology, National Chengchi University, Taipei, Taiwan

[5]Center for Infection Control and Department of Internal Medicine, National Cheng Kung University Hospital, Tainan, Taiwan

[6]Institute of Health Behavior and Community Sciences, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Carol Strong, PhD
Department of Public Health
College of Medicine
National Cheng Kung University
1 University Road
Tainan, 70101
Taiwan
Phone: 886 662353535
Email: carol.chiajung@gmail.com

Abstract

Background: New innovative technologies, such as mobile apps, have been developed to increase pre-exposure prophylaxis (PrEP) adherence and the use of log sex diaries. The contiguity of mobile apps reduces the recall bias that generally affects reported condom and PrEP use. However, none of the currently used mobile apps were designed for event-driven PrEP users, and few studies have demonstrated the potential usage of sex diary data to facilitate the understanding of the different HIV risks among heterogeneous profiles of sex diaries and PrEP use.

Objective: We aim to discriminate the heterogeneous profiles of sex events and PrEP use and examine the risk of condomless anal sex among different types of sex events.

Methods: We recruited 35 adult men who have sex with men from two medical centers in Taiwan since May 2020 and followed up for four months. Participants were on PrEP or willing to take PrEP. They were asked to log their sex events, PrEP use, and dosing regimens on a mobile app to improve their PrEP adherence. Latent class analysis was used to distinguish profiles of sex events and PrEP use. Indicators included correct intake of PrEP for each sex event, participants' sexual positioning, partner's HIV status, and age.

Results: A total of 551 sex events were classified into three classes by latent class analysis: PrEP nonadherent flip-flopping (234/551, 42%), PrEP imperfect-adherent power bottoming (284/551, 52%), and PrEP adherent serodiscordant topping (33/551, 6%). "PrEP nonadherent flip-flopping" sex events were more likely to involve condomless anal sex than "PrEP imperfect-adherent power bottoming" (OR 1.83, 95% CI 1.03-3.25) after considering random intercepts for individuals, and this class needed to increase their PrEP adherence and use of condoms. "PrEP imperfect-adherent power bottoming" realized their own risk and packaged PrEP with condoms to protect themselves. Up to 99% (32/33) of sex events in "PrEP adherent serodiscordant topping" were protected by PrEP, but all of the sex events in this group were condomless.

Conclusions: Using the sex diary data could advance the capacity to identify high-risk groups. HIV prevention strategy should be more flexible and combine PrEP with condom use for future HIV prevention.

(J Med Internet Res 2021;23(12):e33877) doi:10.2196/33877

KEYWORDS

mobile apps; pre-exposure prophylaxis; PrEP; latent class analysis; men who have sex with men; MSM; condom; sex diaries; adherence; app; sex; diary; sexual health; HIV; Taiwan; risk; prevention

Introduction

Pre-exposure prophylaxis (PrEP) is an effective tool for HIV prevention [1]. The number of PrEP users has increased in recent years [2]. In Taiwan, the majority of PrEP users are men who have sex with men (MSM) [3], and it is estimated that 8.9% of MSM were on PrEP in 2019 [4]. A question has emerged from the scale-up of PrEP use regarding whether MSM have decreased their intention of using condoms since they are on PrEP. MSM PrEP users may tend to practice condomless anal sex because they perceive PrEP use decreases their risk of HIV infection [5-8]. A few studies did not find support for the increase of condomless anal sex after PrEP initiation [1,9,10]. The evidence of the association between condomless sex and PrEP use remains inconsistent. The practice of condom use needs to be addressed in PrEP implementation because other sexually transmitted infections may increase. Moreover, if PrEP users are stigmatized as those who prefer condomless sex [11], PrEP scalability may be impeded.

To address whether PrEP use is associated with condomless anal sex in MSM requires measurement of condom use during sex and PrEP intake. Most studies have used self-reported surveys to measure condom use and PrEP adherence [12-15], but recall bias affected the accuracy when participants were asked to recall how many times they used condoms or took PrEP over a long period. Some studies asked participants whether they used condoms during the last anal sex event to reduce the recall bias [16-18], but a one-time sex event neither reflects the long-term decision-making of condom use nor assesses the longitudinal change of PrEP adherence. Although self-reporting bias is still inevitable, using a website survey or mobile apps to measure condom and PrEP use can be more immediate than traditional surveys [19-22]. Additionally, PrEP adherence could be measured by drug concentrations in hair or dried blood spots and Wisepill bottle openings [23]. However, drug concentration testing is costly and time-consuming, and Wisepill bottle openings may underestimate PrEP adherence if the bottle is not being used [23]. There has been no perfect measurement for condom and PrEP use until now, but a mobile app that allows users to log their PrEP use and sex diary at any time or place may be more suitable for monitoring HIV risk in MSM. Sex diaries have been used to track sexual encounters, reduce risky sexual behavior [19-21], and facilitate PrEP uptake [19]. Therefore, a sex diary on a mobile app that gathers the information of each sex event, such as condom use and partner's HIV status, is important for researchers to examine condom use with different sexual partners and track whether or not PrEP is taken correctly.

Because the data from sex diaries contain various characteristics, latent class analysis has usually been used for classifying the characteristics of sexual partners in each sex event to reduce the dimensions of data and create profiles of sexual activities to facilitate outcome analysis. Classifying the patterns of sex events and PrEP use is crucial to identifying which patterns were associated with condomless anal sex. Factors such as a partner's HIV status and age, sex position, condom use, and dosing regimen of PrEP should be considered in relation to whether one has adhered to PrEP as those factors affect the decision-making of condom use. For example, having sex with an older partner was more likely to involve condomless anal sex [24]. MSM who identified as bottoms were less likely to use condoms than those identified as tops [18]. MSM were more likely to practice condomless anal sex with a partner whose HIV status was the same as themselves, compared with those who had sex with a partner whose HIV status was different [25]. Using mobile apps for MSM to log PrEP uptake and sex diaries, including the risk factors mentioned above, can help researchers identify the subgroups of sex events and their relationship with condom use.

In this study, we used event-level data collected from sex diaries in a mobile app that allowed MSM PrEP users to record their PrEP intake, choice of dosing regimen, and information for each sex event. We aimed to understand whether the decision of condom use in a sex event could be evaluated based on one's PrEP use and perception of HIV risk. Specifically, we described the pattern of sex diaries and the proportion of correct PrEP use in relation to condom use and the sexual partners' HIV status. We then used latent class analysis (LCA) to identify heterogeneous sex diaries and PrEP use profiles. Lastly, we examined whether condom use was related to the various sex diaries profiles involving PrEP use. This study demonstrates innovative implications for sex diary data, and the findings will broaden applications in further HIV interventions.

Methods

Study Population and Procedure

We recruited participants referred by physicians and PrEP navigators from two medical centers in two major cities of Taiwan (Taipei and Tainan). The following MSM were eligible for inclusion: (1) HIV-negative, (2) age ≥20 years, (3) currently taking PrEP or willing to initiate PrEP after enrollment, (4) had at least 4 episodes of anal sex with men in the previous month, and (5) are willing to install our mobile app—the UPrEPU app—on their smartphone device. Eligible participants interested in this study were informed about the purpose and goal of the study. In the informed consent process, research assistants explained which types of data would be collected by the UPrEPU app. All the data were de-identified, and participants' personal information was unrecognizable. People who agreed to join the study were asked to provide signed informed consent, and research assistants tutored them regarding the use of the UPrEPU app. Data collected on the app was kept private and confidential. Recruitment began in May 2020, and each participant was followed for 4 months. Participants were tested

for HIV at each monthly follow-up. Participants could receive US $20 cash for each follow-up as an incentive. More detailed study procedures were described elsewhere [26].

The UPrEPU was designed to remind the users to take PrEP based on their dosing regimen, potential time, and pre-arranged date for the sex event and monitor their PrEP adherence. More than half of MSM in Taiwan (56%) chose to take PrEP with an event-driven (ED) dosing regimen [27], a regimen that required individuals to take two pills 2 to 24 hours before sex, and followed by taking one pill 24 hours after sex and another pill 48 hours after sex [28]. This is the first app that considered the complexity of the ED dosing regimen, and the reminders for taking ED PrEP were contingent on the timing of sex events. The UPrEPU app allowed users to track whether their sex events were protected by PrEP and log their dynamic PrEP-dosing choices. Participants could also record relevant information regarding each sexual event, including whether a condom was used, participants' sexual positioning, sexual partner's age, and HIV status. The present study was approved by the Institutional Review Board of the National Cheng Kung University Hospital (IRB: B-BR-107-076-T).

Latent Class Indicators

Correct Intake of Prep for Each Sex Event

For each sex event, whether or the MSM had correctly used PrEP was based on their dosing regimen and categorized as: (1) correct use of the daily dosing regimen, (2) correct use of the ED dosing regimen, and (3) incorrect use of either dosing regimen. Because taking at least 4 pills per week was suggested to provide enough protective PrEP concentrations [29], correct use of a daily dosing regimen was defined as at least 4 pills taken 168 hours before each sex event. Correct use of an ED dosing regimen was defined as taking 2 pills before and another 2 after each sex event. More specifically, the 2 pills before should be taken 2 to 24 hours before each sex event. However, if at least one pill was taken 25 to 168 hours before each sex event, taking only one pill 2 to 24 hours before each sex event was allowed [28]. The selection of when to take the 2 pills after each sex event was based on when the pill was taken right before sex. Our measurement allowed a 2-hour buffer for the participant to log his PrEP use into the app. Participants could also log their previous PrEP use. For example, if participants took 1 or 2 pills X hours before each sex event (X ranged from 2 to 24), they should take another single pill from 22 to 26 hours after the time X. The last pill should be taken 46 to 50 hours after the time X. Incorrect use referred to those in neither of the above groups.

Participants' Sexual Positioning

Participants were asked to report whether they practiced insertive, receptive anal sex, or both during each sex event.

Partner's HIV Status

Response options of each participant's sex partner's HIV status consisted of the following: unknown HIV status, negative and on PrEP, negative but not on PrEP, positive with an undetectable viral load (UVL), or positive with an unknown viral load.

Partner's Age

We asked participants to estimate the age of their sexual partner for each event; the options included less than 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, and above 60 years old. The comparison between the participants' and their sexual partners' age was used as the latent class indicator, including "partner was at least one age category younger than I," "the same as participant's age category," and "partner was at least one age category older than I."

Statistical Analysis

The pattern of sex events during 4 months was presented for each participant. For each sex event, we measured whether a condom was used and whether PrEP was taken correctly, along with the choice of the dosing regimen. LCA was conducted using R version 4.0.3 (R Core Team) and the poLCA package [30]. We used LCA to classify sex events into subgroups. Latent class indicators included participants' PrEP dosing regimens during each sex event, their sexual positions, their sexual partner's HIV status, and the age comparison between participants and their sexual partner. The models were estimated 10 times, with 500 iterations for each run; the solutions with the largest likelihood values were considered maximum likelihood estimates. The model with the lowest Akaike Information Criteria (AIC) [31], lowest Bayesian Information Criteria (BIC) [32], and highest entropy was considered the best-fitting model [33].

Posterior class membership probabilities were used to assign a predicted class to each sex event. Logistic regression analysis was used to compare the difference in condom use between classes. Since one participant could provide more than one sex event, we further used mixed-effects logistic regression with random intercepts for participants.

Results

Participants

The study enrolled 35 users who installed the UPrEPU app. We excluded 3 (8.6%) users who did not enter any sex event in the app. The average age of the remaining 32 (91.4%) users was 29.3 years (SD 4.9). During 4 months, no seroconversion was reported. The number of sex events in 4 months ranged from 1 to 66 (mean number: 17.2). A total of 551 sex events were recorded in this study. Table 1 lists the characteristics of sexual partners and sex events. Condoms were used in 22% (123/551) of sex events. Participants reported correct use of ED PrEP in 64% (352/551) of sex events and correct use of daily PrEP in 14% (79/551). Insertive anal sex was practiced in 35% (193/551) of sex events; receptive anal sex was practiced in 54% (295/551); 11% (63/551) of sex events reported practicing both insertive and receptive anal sex in the same sex event. Among the HIV and PrEP status of users' sexual partners in all events, 38% (209/551) of sexual partners' HIV status was unknown, 21% (118/551) were on PrEP, 30% were negative and not on PrEP, 11% (59/551) were positive with UVL, and there were no positive partners with unknown viral load. Half of the sexual partners (273/551) were reported in the same age category as the participants.

Table 1. Characteristics of sexual partners and sex events (N=551).

Participant characteristics	Values, n (%)
Whether condom was used	
Yes	123 (22)
No	428 (78)
Participants' PrEP[a] dosing regimens during the sex event	
Correct use of daily PrEP	79 (14)
Correct use of ED[b] PrEP	352 (64)
Incorrect use of PrEP	120 (22)
Participants' sexual positioning	
Insertive anal sex	193 (35)
Receptive anal sex	295 (54)
Both insertive and receptive anal sex in the same sex event	63 (11)
Partner's HIV status	
Unknown HIV status	209 (38)
Negative and on PrEP	118 (21)
Negative but not on PrEP	165 (30)
Positive with an undetectable viral load	59 (11)
Positive with an unknown viral load	0 (0)
Partner's age[c]	
Partner was at least one age category younger	154 (28)
The same as participant's age category	273 (50)
Partner was at least one age category older	124 (23)

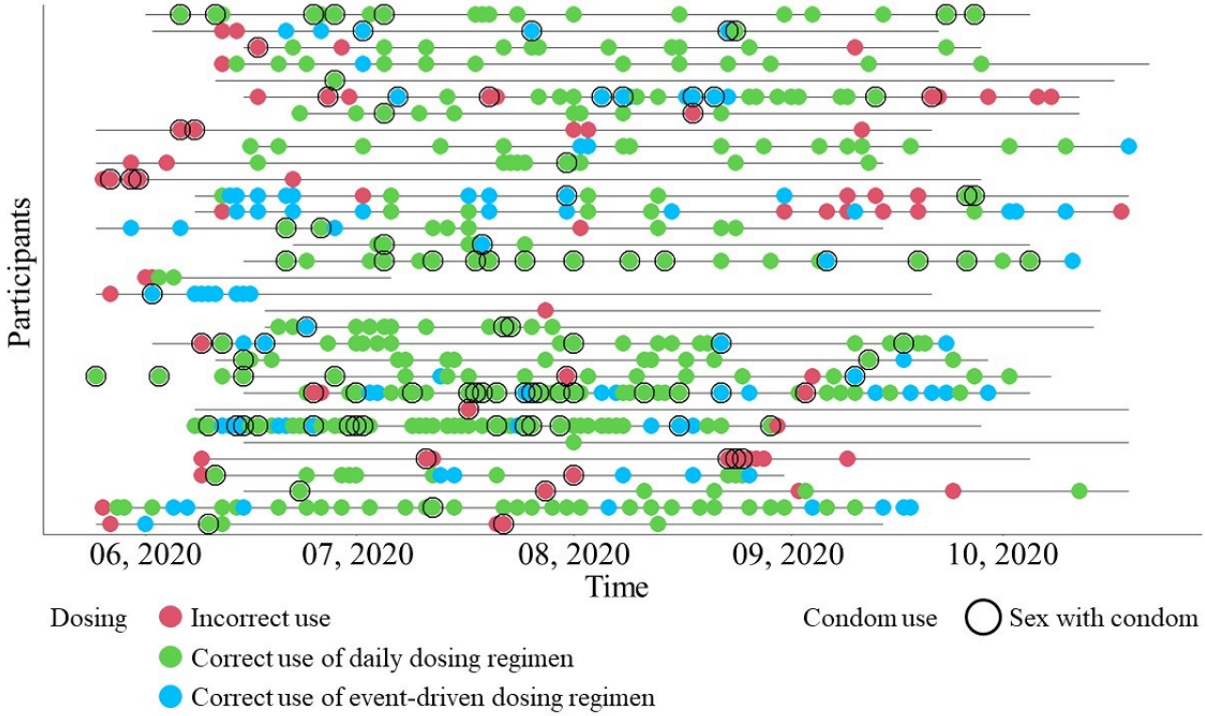
[a]PrEP: pre-exposure prophylaxis.

[b]ED: event-driven.

[c]Age category difference was based on the following age categorical variables of the sexual partners reported by the participants: less than 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, and above 60 years old.

Among 551 sex events, 83% (455/551) were protected by either PrEP or condoms, 18% (99/551) were protected by both PrEP and condoms, and 17% (96/551) were protected neither by PrEP nor by condoms (Figure 1). Among 32 participants, only one (3.1%) person was protected by PrEP and wore condoms during every sex event. There were 7 (22%) MSM protected by PrEP every time that had at least 1 condomless sex event and 24 (75%) MSM that had at least 1 sex event protected neither by PrEP nor a condom.

Figure 1. Sex events of each participant with dosing regimen and condom use (N=551).



Latent Class Analysis

The statistics of one-class to five-class models are shown in Table 2. Compared with a four-class model, the three-class model had lower BIC. Although the four-class model had lower AIC than three clusters, BIC is more appropriate to indicate the best-fitting model [34]. The three-class model also had the highest entropy. Therefore, we selected the three-class model as the best solution.

Table 2. Tests of model fit to identify the optimal number of latent classes.

	Log likelihood	AIC[a]	BIC[b]	Entropy
1 Class	–2306	4629	4668	1.00
2 Class	–2251	4540	4622	0.56
3 Class	–2213	4484	4609	0.76
4 Class	–2187	4453	4621	0.71
5 Class	–2177	4452	4664	0.77

[a]AIC: Akaike Information Criteria.

[b]BIC: Bayesian Information Criteria.

Characterization of Latent Classes

The conditional probabilities of sex events within each class are listed in Table 3. Class 1 comprised 42% (234/551) of the sex events and was named "PrEP nonadherent flip-flopping," showing the highest proportion of incorrect use of PrEP (45/234, 19.2%), performing both insertive and receptive anal sex in the same sex event (58/234, 24.8%), HIV status of partner unknown 103/234, 44.0%), and partner on PrEP (86/234, 36.8%). This class showed the lowest proportion of having a partner at least one age category older (21/234, 9%). Class 2 had the highest proportion of performing receptive anal sex in sex events (281/284, 98.9%), HIV-negative partners, but not on PrEP (130/584, 45.8%). PrEP protected almost 90% (251/284) of sex events in Class 2. Class 2 comprised 52% (284/551) of sex events and was named "PrEP imperfect-adherent power bottoming." Class 3 had the highest proportion of correct use of ED PrEP (29/33, 88.0%), HIV-positive partner with UVL (31/33, 94.0%), performing insertive anal sex (33/33, 100%), and having a partner who was at least one age category older (24/33, 72.7%). It also had the lowest proportion of incorrect use of PrEP (1/33, 3.0%), HIV status of partner unknown (2/33, 6.1%), and having a partner who was in the same age category as the participant (0/33, 0%). Class 3 comprised 6% (33/551) of sex events and was named "PrEP adherent serodiscordant topping."

Table 3. Latent classes and conditional probabilities of sex events (N=551).

	PrEP nonadherent flip-flopping (n=234)	PrEP imperfect-adherent power bottoming (n=284)	PrEP adherent serodiscordant topping (n=33)
Participants' PrEP[a] dosing regimens during the sex event			
Correct use of daily PrEP	0.285	0.165	0.120
Correct use of ED[b] PrEP	0.524	0.720	0.867
Incorrect use of PrEP	0.191	0.114	0.013
Participants' sexual positioning			
Insertive anal sex	0.613	0.011	1.000
Receptive anal sex	0.139	0.989	0.000
Both insertive and receptive anal sex in the same sex event	0.248	0.000	0.000
Partner's HIV status			
Unknown HIV status	0.441	0.363	0.048
Negative and on PrEP	0.368	0.094	0.000
Negative but not on PrEP	0.177	0.457	0.000
Positive with an undetectable viral load	0.014	0.086	0.952
Partner's age[c]			
Partner was at least one age category older	0.090	0.288	0.746
The same as participant's age category	0.563	0.495	0.000
Partner was at least one age category younger	0.347	0.217	0.254

[a]PrEP: pre-exposure prophylaxis.

[b]ED: event-driven.

[c]Age category difference was based on the following age categorical variables of the sexual partners reported by the participants: less than 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, and above 60 years old.

Clusters and Condomless Anal Sex

The proportion of condomless anal sex in each class was 85% (198/234), 69% (197/284), and 100% (33/33), respectively (Table 4). Since the sex events in "PrEP adherent serodiscordant topping" were all condomless, we only compared the probability of condomless anal sex between "PrEP nonadherent flip-flopping" and "PrEP imperfect-adherent power bottoming."

In logistic regression, "PrEP nonadherent flip-flopping" events were 1.43 times more likely to be condomless anal sex compared to "PrEP imperfect-adherent power bottoming" (OR 2.43, 95% CI 1.57-3.76; $P<.001$). After adjusting for the random intercepts, the odds for condomless anal sex were still significantly different between "PrEP nonadherent flip-flopping" and "PrEP imperfect-adherent power bottoming" (OR 1.83, 95% CI 1.03-3.25; $P=.04$).

Table 4. Logistic regression for condomless anal intercourse among three classes.

	% of condomless anal sex	OR	95% CI	*P* value
Model 1				
PrEP[a] nonadherent flip-flopping	85%	2.43	1.57-3.76	<.001
ref: PrEP imperfect-adherent power bottoming[b]	69%	—	—	—
PrEP adherent serodiscordant topping[c]	100%	—	—	—
Model 2				
PrEP nonadherent flip-flopping[d]	85%	1.83	1.03-3.25	.04
ref: PrEP imperfect-adherent power bottoming[b]	69%	—	—	—
PrEP adherent serodiscordant topping[c]	100%	—	—	—

[a]PrEP: pre-exposure prophylaxis.

[b]PrEP imperfect-adherent power bottoming was the reference group; hence, empty cells were shown in the table.

[c]PrEP adherent serodiscordant topping was not included in logistic regression due to 100% of condomless anal intercourse; hence empty cells were shown in the table.

[d]Adding random intercepts for individuals.

Discussion

Our study outlined 3 types of users and their behavioral patterns through analyzing their logs. This is the first study to classify event-level data and investigate the relationship between condom use and the heterogeneous profiles of sexual activities, including PrEP uptake among MSM. We demonstrated the use of event-level data of PrEP intake and sex diary logs from a mobile app to capture PrEP adherence in the real world. The sex diary allowed users to self-monitor and researchers to track unprotected sexual behavior and PrEP adherence. Even though high PrEP adherence was observed in our sample, only one-quarter (8/32) of participants had 100% PrEP adherence during the study period. If other HIV prevention methods such as condoms were practiced, imperfect PrEP adherence would not always be a concern; however, if neither PrEP nor condom use is well-adhered, such sexual activities would be exposed to higher HIV risk.

For example, “PrEP nonadherent flip-flopping” required more public health attention because they were neither PrEP nor condom protected. “PrEP nonadherent flip-flopping” comprised both bottoms and tops on a mixed PrEP-dosing regimen with unknown HIV status partners and had significantly lower odds of using condoms versus the “PrEP imperfect-adherent power bottoming” group. One possible reason for not using condoms among the “PrEP nonadherent flip-flopping” group may be that up to one-third (86/234, 36.8%) of sexual partners in this group were HIV-negative and on PrEP. The probability of HIV infection from a partner who was on PrEP was relatively low, and it may have resulted in reduced intention to use condoms [5,6]; however, 44% (103/234) of partners were still HIV-unknown in this group. Even though no participants acquired HIV in the 4-month study period, this group may have been exposed to HIV risk longitudinally. MSM require interventions to assist in taking PrEP correctly, such as reminders or notifications from mobile apps or individualized consultations with their PrEP navigators about their difficulties.

We identified one class, the “PrEP imperfect-adherent power bottoming,” mainly comprised of bottoms with HIV-negative partners, not on PrEP. This group may have been aware of their own risk and therefore practiced HIV prevention behavior: 90% (251/284) of sex events were protected by PrEP, and almost one-third (87/284, 30.6%) were protected by condoms. A discrete choice experiment study used conditional logic analysis to understand the preference for PrEP and condoms based on the hypothesized risk of HIV among gay, bisexual, and other MSM and showed that condoms were used as additional protection besides PrEP when there is a perceived increase in HIV risk [35], which was similar to the behavior of condom use in the “PrEP imperfect-adherent power bottoming.” Whether MSM in this class used condoms depended on their perception of the HIV risk in each sex event. This was also the largest class in our sample, which showed a potential for a combination prevention intervention that adapts to the fluctuating risk for each individual.

The “PrEP adherent serodiscordant topping” class, mainly comprised of tops and partners who were HIV positive with UVL, chose to protect themselves with PrEP instead of condoms—all sex events in this class were condomless. Reasons for MSM to have sex without condoms if they were already on PrEP include reduced fear of HIV, lessened anxiety of having sex due to PrEP use, and maintaining sexual pleasure [6,7]. Furthermore, MSM might decide not to use condoms when having sex with a regular sexual partner due to the mutual trust that friends would not put them at risk [6]. Some MSM indicated that they did not want to use condoms, and therefore they showed strong adherence to PrEP [6]. In this class, we found that PrEP had been successfully implemented, and the concept of “undetectable equals untransmissible” had been well-accepted. Public health researchers still need to stress the importance of using condoms to prevent other sexually transmitted infections. However, with the broadened options for HIV prevention, condoms might not be the first choice for many people in the real world. Therefore, health educational

campaigns may need to reprioritize prevention information to achieve optimal effectiveness.

Sex diary data used in this study uniquely included measurement for ED PrEP adherence, such as hours of sex events, instead of only dates. Compared to the apps designed for a daily dosing regimen, the UPrEPU app provided reminders for ED PrEP users to take PrEP based on when sex happened and users' previous dosing regimens. Adherence to ED PrEP can be calculated only when the hours of the sex event can be logged on an app such as UPrEPU. Such a feature is particularly important since ED PrEP use is more prevalent during the COVID-19 pandemic [36].

The strength of our study was its use of sex diaries to capture the fluctuating HIV risk of sexual behaviors and to classify the diverse patterns of sex events. Most studies have identified the different HIV risks based on the individual level [12-15]. They categorized MSM into different subtypes, but MSM perform various types of sexual behaviors in the real world, which suggests that analyzing event-level data is closer to reality. Identifying the risky class from the diverse patterns of sex events helps researchers build up a specific intervention more intentionally. We offer a more comprehensive recommendation for MSM by looking at different sex roles and PrEP adherence. Based on our findings, future mobile apps can consider giving notifications according to each user's previous patterns of sex events.

Limitations

This study has the following limitations. First, participants could see whether they took PrEP correctly in each sex event from the UPrEPU app. This may have interfered with their own judgment for whether they were protected by PrEP and affected their decision to use condoms. Without the app's assistance, PrEP users might not perceive the PrEP adherence correctly, and the decision-making of condom use would be relying solely on their self-perception, which could either be over-estimating or underestimating PrEP adherence.

Second, some contexts of sexual activities were not yet collected in our study. More details of sexual activities may change the categorization in LCA. Information such as whether the partner is a steady sex partner, uses chemsex/sexualized drugs, and belongs to particular sexual networks may be confounders for the association between PrEP adherence and condom use. Lastly, the representative of the study population may be limited due to the recruitment and inclusion criteria. We included participants from only two medical centers and those who had a relatively high frequency of sex events, which restricted the generalizability to MSM who did not use PrEP to protect themselves or whose frequency of sex events was low.

Conclusions

Our study advances the understanding of the association between HIV risk and the different types of sexual activities. We demonstrated the potential to use mobile app-based sex and PrEP diary for the future development of effective intervention programs by identifying high-risk groups. This study identified that one class, "PrEP nonadherent flip-flopping," demanded more support for implementing an HIV prevention strategy. "PrEP imperfect-adherent power bottoming" demonstrated that condoms were used as extra protection other than PrEP, and "PrEP adherent serodiscordant topping" showed almost perfect PrEP adherence but risked other sexually transmitted infections. Future HIV interventions should consider combining various strategies to improve the effectiveness of HIV prevention programs.

Acknowledgments

This study was supported by the Taiwan Ministry of Science and Technology (MOST 108-2636-B-006-004, 109-2636-B-006-004, and MOST 110-2636-B-006-011) and the Taiwan Ministry of Health and Welfare, Centers for Disease Control (MOHW 109-CDC-C-114-000104). All authors are very appreciative of all participants involved in the pilot study of the UPrEPU app. The MOST and the Taiwan Centers for Disease Control had no role in the design of the study and the collection, analysis, or interpretation of data or writing of the manuscript.

Authors' Contributions

CS, HW, and YY conceived, designed, and performed the study. HW and YY analyzed the data and drew the figure and tables. YY and CS wrote the manuscript. YY, CS, HW, SK, PH, CL, and PH revised the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Fonner V. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. AIDS (London, England) 2016 Jul 31;30(12):1973-1983. [doi: 10.1097/qad.0000000000001145]
2. MacGibbon J, Lea T, Ellard J, Murphy D, Kolstee J, Power C, et al. Access to Subsidized Health Care Affects HIV Pre-Exposure Prophylaxis (PrEP) Uptake Among Gay and Bisexual Men in Australia: Results of National Surveys 2013-2019. JAIDS Journal of Acquired Immune Deficiency Syndromes 2021 Apr 01;86(4):430-435. [doi: 10.1097/QAI.0000000000002572] [Medline: 33230031]

3. Wu H, Ku SW, Li C, Ko N, Yu T, Chung A, et al. Factors Associated with Preferred Pre-Exposure Prophylaxis Dosing Regimen Among Men Who Have Sex with Men in Real-World Settings: A Mixed-Effect Model Analysis. AIDS and Behavior 2020 Jul 08;25(1):249-258. [doi: 10.1007/s10461-020-02964-5]
4. Ogunbajo A, Kang A, Shangani S, Wade RM, Onyango DP, Odero WW, et al. Awareness and Acceptability of pre-exposure prophylaxis (PrEP) among gay, bisexual and other men who have sex with men (GBMSM) in Kenya. AIDS Care 2019 Apr 30;31(10):1185-1192. [doi: 10.1080/09540121.2019.1612023]
5. Carlo Hojilla J, Koester KA, Cohen SE, Buchbinder S, Ladzekpo D, Matheson T, et al. Sexual Behavior, Risk Compensation, and HIV Prevention Strategies Among Participants in the San Francisco PrEP Demonstration Project: A Qualitative Analysis of Counseling Notes. AIDS and Behavior 2015 Apr 3;20(7):1461-1469. [doi: 10.1007/s10461-015-1055-5]
6. Gafos M, Horne R, Nutland W, Bell G, Rae C, Wayal S, et al. The Context of Sexual Risk Behaviour Among Men Who Have Sex with Men Seeking PrEP, and the Impact of PrEP on Sexual Behaviour. AIDS and Behavior 2018 Oct 10;23(7):1708-1720. [doi: 10.1007/s10461-018-2300-5]
7. Storholm ED, Volk JE, Marcus JL, Silverberg MJ, Satre DD. Risk Perception, Sexual Behaviors, and PrEP Adherence Among Substance-Using Men Who Have Sex with Men: a Qualitative Study. Prevention Science 2017 Jun 3;18(6):737-747. [doi: 10.1007/s11121-017-0799-8]
8. Holt M, Lea T, Mao L, Kolstee J, Zablotska I, Duck T, et al. Community-level changes in condom use and uptake of HIV pre-exposure prophylaxis by gay and bisexual men in Melbourne and Sydney, Australia: results of repeated behavioural surveillance in 2013–17. The Lancet HIV 2018 Aug 05;5(8):e448-e456. [doi: 10.1016/s2352-3018(18)30072-9]
9. Sagaon-Teyssier L, Suzan-Monti M, Demoulin B, Capitant C, Lorente N, Préau M, for the ANRS IPERGAY Study Group. Uptake of PrEP and condom and sexual risk behavior among MSM during the ANRS IPERGAY trial. AIDS Care 2016 Feb 17;28(sup1):48-55. [doi: 10.1080/09540121.2016.1146653]
10. Mugwanya KK, Donnell D, Celum C, Thomas KK, Ndase P, Mugo N, et al. Sexual behaviour of heterosexual men and women receiving antiretroviral pre-exposure prophylaxis for HIV prevention: a longitudinal analysis. The Lancet Infectious Diseases 2013 Oct 17;13(12):1021-1028. [doi: 10.1016/s1473-3099(13)70226-3]
11. Franks J, Hirsch-Moverman Y, Loquere AS, Amico KR, Grant RM, Dye BJ, et al. Sex, PrEP, and Stigma: Experiences with HIV Pre-exposure Prophylaxis Among New York City MSM Participating in the HPTN 067/ADAPT Study. AIDS and Behavior 2017 Nov 15;22(4):1139-1149. [doi: 10.1007/s10461-017-1964-6]
12. Dangerfield DT, Carmack CC, Gilreath TD, Duncan DT. Latent Classes of Sexual Positioning Practices and Sexual Risk Among Men Who Have Sex with Men in Paris, France. AIDS and Behavior 2018 Sep 1;22(12):4001-4008. [doi: 10.1007/s10461-018-2267-2]
13. Kelly JA, DiFranceisco WJ, St. Lawrence JS, Amirkhanian YA, Anderson-Lamb M. Situational, Partner, and Contextual Factors Associated with Level of Risk at Most Recent Intercourse Among Black Men Who Have Sex with Men. AIDS and Behavior 2013 Jul 19;18(1):26-35. [doi: 10.1007/s10461-013-0532-y]
14. Achterbergh RCA, de Vries HJC, Boyd A, Davidovich U, Drückler S, Hoornenborg E, et al. Identification and characterization of latent classes based on drug use among men who have sex with men at risk of sexually transmitted infections in Amsterdam, the Netherlands. Addiction 2020 Jan 07;115(1):121-133. [doi: 10.1111/add.14774] [Medline: 31400174]
15. Chan PA, Rose J, Maher J, Benben S, Pfeiffer K, Almonte A, et al. A Latent Class Analysis of Risk Factors for Acquiring HIV Among Men Who Have Sex with Men: Implications for Implementing Pre-Exposure Prophylaxis Programs. AIDS Patient Care and STDs 2015 Oct 27;29(11):597-605. [doi: 10.1089/apc.2015.0113]
16. Schnarrs PW, Rosenberger JG, Schick V, Delgado A, Briggs L, Dodge B, et al. Difference in Condom Use Between Bear Concordant and Discordant Dyads During the Last Anal Sex Event. Journal of Homosexuality 2016 May 13;64(2):195-208. [doi: 10.1080/00918369.2016.1174024]
17. Sandfort T, Yi H, Knox J, Reddy V. Sexual Partnership Types as Determinant of HIV Risk in South African MSM: An Event-Level Cluster Analysis. AIDS and Behavior 2012 Sep 6;17(S1):23-32. [doi: 10.1007/s10461-012-0294-y]
18. Dangerfield DT, Gravitt P, Rompalo AM, Tai R, Lim SH. Correlates of anal sex roles among Malay and Chinese MSM in Kuala Lumpur, Malaysia. International journal of STD & AIDS 2015 Apr 16;27(4):313-320. [doi: 10.1177/0956462415581125]
19. Liu AY, Laborde ND, Coleman K, Vittinghoff E, Gonzalez R, Wilde G, et al. DOT Diary: Developing a Novel Mobile App Using Artificial Intelligence and an Electronic Sexual Diary to Measure and Support PrEP Adherence Among Young Men Who Have Sex with Men. AIDS and Behavior 2020 Oct 12;25(4):1001-1012. [doi: 10.1007/s10461-020-03054-2]
20. Finkenflügel RN, Hoornenborg E, Achterbergh RC, Marra E, Davidovich U, de Vries HJ, et al. A Mobile Application to Collect Daily Data on Preexposure Prophylaxis Adherence and Sexual Behavior Among Men Who Have Sex With Men: Use Over Time and Comparability With Conventional Data Collection. Sexually transmitted diseases 2019 Jun;46(6):400-406. [doi: 10.1097/olq.0000000000000999]
21. Songtaweesin WN, Kawichai S, Phanuphak N, Cressey TR, Wongharn P, Saisaengjan C, CE-PID - TRC Adolescent Study Team. Youth-friendly services and a mobile phone application to promote adherence to pre-exposure prophylaxis among adolescent men who have sex with men and transgender women at-risk for HIV in Thailand: a randomized control trial. Journal of the International AIDS Society 2020 Sep 31;23 Suppl 5(S5):e25564 [FREE Full text] [doi: 10.1002/jia2.25564] [Medline: 32869511]

22. Biello KB, Hill-Rorie J, Valente PK, Futterman D, Sullivan PS, Hightow-Weidman L, et al. Development and Evaluation of a Mobile App Designed to Increase HIV Testing and Pre-exposure Prophylaxis Use Among Young Men Who Have Sex With Men in the United States: Open Pilot Trial. Journal of Medical Internet Research 2021 Mar 24;23(3):e25107 [FREE Full text] [doi: 10.2196/25107] [Medline: 33759792]
23. Koss CA, Hosek SG, Bacchetti P, Anderson PL, Liu AY, Horng H, et al. Comparison of Measures of Adherence to Human Immunodeficiency Virus Preexposure Prophylaxis Among Adolescent and Young Men Who Have Sex With Men in the United States. Clinical Infectious Diseases 2018 Jan 06;66(2):213-219 [FREE Full text] [doi: 10.1093/cid/cix755] [Medline: 29020194]
24. Miller RL, Forney JC, McNall MA. Recent sexual partnerships among adolescent and emerging adult black men who have sex with men: The role of age and race discordant partnerships in risk-taking behavior. Journal of HIV/AIDS & Social Services 2014 Aug 21;13(3):252-270. [doi: 10.1080/15381501.2012.748584]
25. Van der Bij AK, Kolader ME, de Vries HJC, Prins M, Coutinho RA, Dukers NHTM. Condom use rather than serosorting explains differences in HIV incidence among men who have sex with men. JAIDS Journal of Acquired Immune Deficiency Syndromes 2007 Aug 15;45(5):574-580. [doi: 10.1097/qai.0b013e3180959ab7]
26. Strong C, Wu H, Tseng Y, Yuan C, Yu Y, Liao JC, et al. Mobile App (UPrEPU) to Monitor Adherence to Pre-exposure Prophylaxis in Men Who Have Sex With Men: Protocol for a User-Centered Approach to Mobile App Design and Development. JMIR Research Protocols 2020 Dec 1;9(12):e20360. [doi: 10.2196/20360]
27. Wu H, Wen-Wei Ku S, Chang HH, Li C, Ko N, Strong C. Imperfect adherence in real life: a prevention-effective perspective on adherence to daily and event-driven HIV pre-exposure prophylaxis among men who have sex with men - a prospective cohort study in Taiwan. Journal of the International AIDS Society 2021 May 20;24(5):e25733 [FREE Full text] [doi: 10.1002/jia2.25733] [Medline: 34018330]
28. Yen-Hao Chu I, Wen-Wei Ku S, Li C, Toh HS, Yang C, Wu K, et al. Taiwan guideline on oral pre-exposure prophylaxis for HIV prevention?2018 update. Journal of Microbiology, Immunology and Infection 2020 Feb;53(1):1-10. [doi: 10.1016/j.jmii.2019.09.003]
29. Vaccher SJ, Marzinke MA, Templeton DJ, Haire BG, Ryder N, McNulty A, et al. Predictors of Daily Adherence to HIV Pre-exposure Prophylaxis in Gay/Bisexual Men in the PRELUDE Demonstration Project. AIDS and Behavior 2018 Dec 06;23(5):1287-1296. [doi: 10.1007/s10461-018-2353-5]
30. Linzer DA, Lewis JB. poLCA: An R Package for Polytomous Variable Latent Class Analysis. Journal of statistical software 2011 Jun;42(10):1-29. [doi: 10.18637/jss.v042.i10]
31. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. Selected Papers of Hirotugu Akaike 1998:199-213. [doi: 10.1007/978-1-4612-1694-0_15]
32. Schwarz G. Estimating the Dimension of a Model. Annals of statistic 1978 Mar 1;6(2):461-464. [doi: 10.1214/aos/1176344136]
33. Larose C, Harel O, Kordas K, Dey DK. Latent class analysis of incomplete data via an entropy-based criterion. Statistical Methodology 2016 Sep;32:107-121. [doi: 10.1016/j.stamet.2016.04.004]
34. Yang C. Evaluating latent class analysis models in qualitative phenotype identification. Computational Statistics & Data Analysis 2006 Feb;50(4):1090-1104. [doi: 10.1016/j.csda.2004.11.004]
35. Tan RKJ, Wang Y, Prem K, Harrison-Quintana J, Teo AKJ, Kaur N, et al. HIV Pre-Exposure Prophylaxis, Condoms, or Both? Insights on Risk Compensation Through a Discrete Choice Experiment and Latent Class Analysis Among Men Who Have Sex With Men. Value in Health 2021 May;24(5):714-723. [doi: 10.1016/j.jval.2020.11.023]
36. Chow E, Hocking JS, Ong JJ, Schmidt T, Buchanan A, Rodriguez E, et al. Changing the Use of HIV Pre-exposure Prophylaxis Among Men Who Have Sex With Men During the COVID-19 Pandemic in Melbourne, Australia. Open Forum Infect Dis 2020 Jul;7(7):ofaa275 [FREE Full text] [doi: 10.1093/ofid/ofaa275] [Medline: 32704518]

Abbreviations

AIC: Akaike Information Criteria
BIC: Bayesian Information Criteria
ED: event-driven
LCA: latent class analysis
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
UVL: undetectable viral load

Edited by G Eysenbach; submitted 27.09.21; peer-reviewed by TH Kwan, B Ma; comments to author 18.10.21; revised version received 26.10.21; accepted 21.11.21; published 23.12.21.

Please cite as:
Yu YF, Wu HJ, Ku SWW, Huang PH, Li CW, Huang P, Strong C
Condomless Anal Sex Associated With Heterogeneous Profiles Of HIV Pre-Exposure Prophylaxis Use and Sexual Activities Among Men Who Have Sex With Men: A Latent Class Analysis Using Sex Diary Data on a Mobile App
J Med Internet Res 2021;23(12):e33877
URL: https://www.jmir.org/2021/12/e33877
doi:10.2196/33877
PMID:34941560

Original Paper

Development of a Supportive Parenting App to Improve Parent and Infant Outcomes in the Perinatal Period: Development Study

Shefaly Shorey[1], PhD; Thiam Chye Tan[2], MMed; thilagamangai[3]; Jancy Mathews[4], MN; Chun Yan Yu[5], MSc; Siew Hoon Lim[6], PhD; Luming Shi[7], MBBS, MSc; Esperanza Debby Ng[1], BPsychSc, BA; Yiong Huak Chan[8], PhD; Evelyn Law[9], MD; Cornelia Chee[9], MBBS, MMed; Yap Seng Chong[9,10], MBBS, MMed, MD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Mount Elizabeth Novena Specialist Centre, Singapore, Singapore
[3]Division of Nursing, KK Women's and Children's Hospital, Singapore, Singapore
[4]National University Polyclinics, Singapore, Singapore
[5]Raffles Medical Group, Singapore, Singapore
[6]Singapore General Hospital, Singapore, Singapore
[7]Singapore Clinical Research Institute, Singapore, Singapore
[8]Biostatistics Unit, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[9]National University Hospital, Singapore, Singapore
[10]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Shefaly Shorey, PhD
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 2, Clinical Research Centre, Block MD11
10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 66011294
Email: nurssh@nus.edu.sg

Abstract

Background: The transition to parenthood can be challenging, and parents are vulnerable to psychological disorders during the perinatal period. This may have adverse long-term consequences on a child's development. Given the rise in technology and parents' preferences for mobile health apps, a supportive mobile health intervention is optimal. However, there is a lack of a theoretical framework and technology-based perinatal educational intervention for couples with healthy infants.

Objective: The aim of this study is to describe the Supportive Parenting App (SPA) development procedure and highlight the challenges and lessons learned.

Methods: The SPA development procedure was guided by the information systems research framework, which emphasizes a nonlinear, iterative, and user-centered process involving 3 research cycles—the relevance cycle, design cycle, and rigor cycle. Treatment fidelity was ensured, and team cohesiveness was maintained using strategies from the Tuckman model of team development.

Results: In the relevance cycle, end-user requirements were identified through focus groups and interviews. In the rigor cycle, the user engagement pyramid and well-established theories (social cognitive theory proposed by Bandura and attachment theory proposed by Bowlby) were used to inform and justify the features of the artifact. In the design cycle, the admin portal was developed using Microsoft Visual Studio 2017, whereas the SPA, which ran on both iOS and Android, was developed using hybrid development tools. The SPA featured knowledge-based content, informational videos and audio clips, a discussion forum, chat groups, and a frequently asked questions and expert advice section. The intervention underwent iterative testing by a small group of new parents and research team members. Qualitative feedback was obtained for further app enhancements before official implementation. Testing revealed user and technological issues, such as web browser and app incompatibility, a lack of notifications for both administrators and users, and limited search engine capability.

Conclusions: The information systems research framework documented the technical details of the SPA but did not take into consideration the interpersonal and real-life challenges. Ineffective communication between the health care research team and the app developers, limited resources, and the COVID-19 pandemic were the main challenges faced during content development. Quick adaptability, team cohesion, and hindsight budgeting are crucial for intervention development. Although the effectiveness of the SPA in improving parental and infant outcomes is currently unknown, this detailed intervention development study highlights the key aspects that need to be considered for future app development.

(J Med Internet Res 2021;23(12):e27033) doi:10.2196/27033

KEYWORDS

depression; development; education; parent; perinatal; support; telehealth; mobile phone

Introduction

Background

In recent years, depression has displaced many chronic diseases to be the leading cause of disability worldwide and the main contributor to the overall global burden of diseases [1]. Although depression can occur at any point in life, women and men are at increased risk of depression in the postpartum period [2,3]. The transition to parenthood, coupled with emotional, physiological, hormonal, and psychosocial changes experienced during childbirth, could have increased women's susceptibility to postpartum depression [4-6]. The psychological state of women could influence their partners' psychological well-being because maternal postpartum depression is a known predictor of paternal postpartum depression, affecting 24%-50% of all fathers [7]. Fathers also have the highest risk of developing paternal postpartum depression 3 to 6 months after childbirth [8,9]. Moreover, research has shown the ripple effects of postpartum depression adversely affecting an infant's subsequent cognitive, physical, social-emotional, and behavioral development over the years [10,11]. Children of mothers who were depressed were found to score lower in intellectual and motor development, and they tend to display insecure attachment, behavioral problems, poor emotional regulation and academic performances, and are at an increased risk of developing psychiatric disorders [12-14]. In addition, a child's internalizing and externalizing behavior problems are significantly related to paternal postpartum depression [15].

Given the potential widespread consequences of postpartum depression on the family unit, there is a strong impetus for couple-based interventions to curb postpartum depression. However, to date, the cause of postpartum depression among parents remains inconclusive, thus inhibiting the development of direct intervention. Instead, research has suggested a multifactorial etiology, as there has been evidence that various psychosocial variables, such as life stress, marital conflict, childcare stress, parenting self-efficacy, poor childcare knowledge, the lack of social support, and antenatal anxiety, are key predictors of postpartum depression [5,6]. Hence, a perinatal intervention targeting these predictors may help to alleviate feelings of stress and depression among parents. Reviews on previous interventions implemented during the perinatal period have indicated the need for more multimodal family-oriented educational programs that focus on parental physical, emotional, and social well-being [16-18]. Furthermore, theory-based educational interventions are scarce, and current interventions are often compartmentalized as antenatal or postpartum [16-19] and focus on specific groups of parents such as those with infants in the neonatal intensive care unit or with behavioral problems [16-18,20,21]. Therefore, there is an urgent need for a comprehensive perinatal educational intervention for parents with healthy infants to enhance parental psychological well-being and infant developmental outcomes.

Value of Technology

Technology-enhanced health care and telemedicine have been gaining popularity because of their better accessibility, flexibility, cost-effectiveness, and ability to provide patient privacy and reduce stigmatization [22,23]. The rapid proliferation in technology-based health care is enabled by the increased use and advancements of mobile apps, wearable devices, live audiovisual communication, and telecommunication. The value of telehealth care and technology-based interventions has been further amplified during the ongoing COVID-19 pandemic as most households worldwide have faced lockdown restrictions, quarantine measures, and social isolation, which limit couples' access to help and support [24]. Moreover, such technology-based interventions are especially crucial in a conservative society such as Singapore's, where traditional views and postpartum practices (eg, home confinement) as well as the stigmatization of mental health disorders restrict parents from seeking help in the postpartum period [25,26].

Previous studies examining the needs of fathers and mothers during the perinatal period have also highlighted parents' preferences for technology-based educational programs, especially mobile health (mHealth) apps [26,27]. mHealth apps have a relatively new edge in health care innovation that allows the delivery of health care and dissemination of educational information [27]. In 2010, more than 200 million mHealth apps were downloaded, and approximately 70% of the population worldwide has access to at least one mHealth app [28]. mHealth pregnancy-related apps are particularly popular among young parents, suggesting a shift in patient empowerment in terms of maternity care [29,30]. However, parents have expressed difficulty in finding pregnancy apps that provide localized and individualized advice and feedback from health care professionals [27,30]. Current literature has also revealed that most of the available mHealth apps are not validated or developed by health care professionals [31], thus calling into question the trustworthiness and credibility of the information provided.

Given the benefits of mHealth apps and the paucity and limitations of current pregnancy-related mHealth apps, this study aims to develop the Supportive Parenting App (SPA), a theory-based perinatal educational mobile app for parents, and to subsequently examine its effectiveness in preventing postpartum depression and enhancing parenting self-efficacy, perceived social support, parent-child bonding, and physical and social-emotional development of an infant. According to the Singapore Infocomm Media Development Authority [32], 98% of households in Singapore in 2019 had internet access and 89% of individuals were internet users. In addition, the smartphone penetration rate of 82% in Singapore is one of the highest in the Asia Pacific region [33]. Therefore, technology is available in most households to support the implementation of the SPA during the perinatal period.

Detailed intervention development studies are limited because of research funding priorities that focused on efficacy or effectiveness studies [34]. However, a comprehensive, systematic, and transparent reporting of intervention development is necessary to help readers understand the challenges and benefits of different intervention development approaches and allow retrospective assessments to be performed to understand how different development approaches may affect the effectiveness of interventions. Therefore, this study aims to provide an insight into the SPA development process and highlight the challenges and lessons faced during the SPA development to improve future mHealth studies.

Project Overview

The SPA project was a multisite study that was conducted in tertiary public hospitals in Singapore. The intervention was delivered to both parents (fathers and mothers), and they were granted access to the SPA during the perinatal period (from the viability age of 24 weeks of gestation to 6 months after childbirth) when the chances of developing antenatal or postpartum depression were very high. The parents were able to access the mobile app anytime and anywhere. The SPA featured educational materials, a discussion forum, chat groups with peer volunteers, and advice from health care professionals. The inclusion criteria were as follows: (1) parents with a female partner who scored ≥9 on the Edinburgh Postnatal Depression Scale, (2) those aged ≥21 years, (3) those who were able to read and speak English, (4) those who had a low-risk pregnancy with more than 24 weeks of gestation (age of viability), and (5) those who had a smartphone with internet access. The exclusion criteria were as follows: parents who (1) had physical or mental disorders that would interfere with their ability to participate in the study; (2) had a high-risk pregnancy, including placenta previa major, preeclampsia, and pregnancy-induced hypertension; or (3) gave birth to a stillborn baby or a newborn with congenital anomalies or medical complications, including pathological jaundice that required special care in hospital. Mothers aged ≥21 years who had previously recovered from postpartum depression were recruited and trained by a psychiatrist to become peer volunteers to befriend and provide adequate web-based support to the parents.

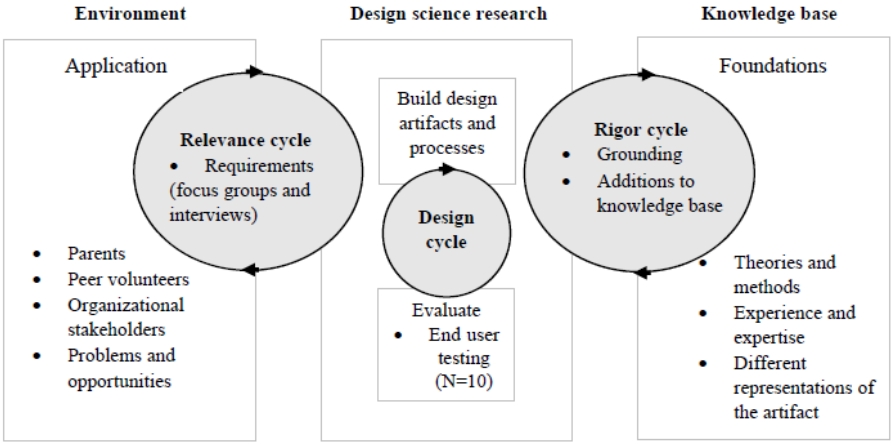
The aims of this research project are to (1) develop a theory-based SPA intervention for parents across the perinatal period; (2) examine the intervention's effectiveness in reducing parental postpartum depression and anxiety and encouraging parenting self-efficacy, help-seeking behavior, parent-infant bonding, parenting satisfaction, and infant developmental outcomes during the perinatal period up to 1 year after childbirth; (3) evaluate the SPA's cost-effectiveness compared with standard perinatal care across major restructured hospitals; and (4) examine SPA user experiences. However, this paper only focuses on the SPA intervention development process, the challenges faced by the research team, and the lessons learned.

Methods

Information Systems Research Framework

The information systems research (ISR) framework, which promotes an iterative, user-centered, nonlinear design process was used to guide the development of the SPA [35]. The ISR framework is commonly used in the creation of technological artifacts and comprises 3 research cycles (relevance cycle, design cycle, and rigor cycle) that link 3 research domains (knowledge base, design science research, and environment). The relevance cycle involves understanding the end users' needs and requirements through feedback and focus groups. The design cycle is where the artifact is developed and evaluated, and the rigor cycle is where theories, past methods, and representations of the artifacts contribute to the current design and development of the new artifact [35]. This study's adaptation of the ISR framework [35] is presented in Figure 1.

Figure 1. Adaptation of the information systems research framework.



Treatment Fidelity

To establish treatment fidelity, the strategies proposed by Bellg et al [36] and Eaton et al [37] for intervention design, training of providers, intervention delivery, and receipt of intervention were adopted. During the intervention design phase, intervention fidelity was ensured by using a theoretical framework (social cognitive theory) as the foundation of the intervention and by preparing a protocol for the intervention. In anticipation of provider and peer volunteer dropout, backup providers and volunteers with the requisite skills were on constant standby. To ensure consistency, delivery of intervention materials (eg, knowledge-based content and expert advice) was standardized by providing briefings and issuing providers and peer volunteers a handbook on the intervention procedures. In terms of receipt of intervention, participants' frequency of access to the app and the use of audio and video materials and peer discussion forum were monitored and recorded through a back-end admin portal. Weekly reminders were sent to parents to reinforce the use of the intervention by accessing the SPA.

Team Development

As the SPA development process involved stakeholders across various departments, immense collaborative effort and a high-performing team were crucial to ensure consistent progress and adherence to the project timeline. The Tuckman model of team development [38] highlighted different phases of team development—forming (initial meeting and beginning of collective work), storming (dealing with confusion and conflict over decision-making and roles), norming (accepting goals, roles, and work positively), performing (focusing on achieving goals and obtaining personal growth), and adjourning (recognizing and celebrating achievements). Although all phases are necessary for successful team building, problem solving, work planning, and delivering of results, the team leader plays a critical role in regulating the team dynamics and ensuring effective workflow by applying appropriate conflict resolution strategies as recommended by Tuckman [38].

Results

Relevance Cycle

The end-user environment and requirements were obtained through a focus group session with new parents. In all, 3 pairs of parents representing each major ethnic group in Singapore (Chinese, Malay, and Indian) were interviewed on their informational and perinatal educational needs and their expectations of the SPA. These parents were not included in the subsequent main study. In addition, because the objectives and components of the SPA were similar to those of the research team's previous research projects—*Home-but-not-Alone* (an mHealth app-based postnatal educational program) and the *Peer-Support Intervention Program*—user requirements were obtained from the projects' focus groups and interviews [39,40]. Previous local research among Singaporean mothers who only received standard routine hospital care indicated a need for more support after childbirth, such as having peer volunteers to talk to, more follow-up appointments with clinicians, and having an educational parenting mobile app catered to parents in Singapore [40]. Mothers who received the intervention requested the inclusion of individualized push notifications to mark important baby developmental milestones and special parenting tips. A study that investigated the use of a technology-based supportive intervention among multiracial Singaporean mothers during the postpartum period suggested the extension of the parenting support programs beyond 1 month after childbirth, introducing the parenting app early in pregnancy, and including a greater variety of educational topics [39].

Rigor Cycle

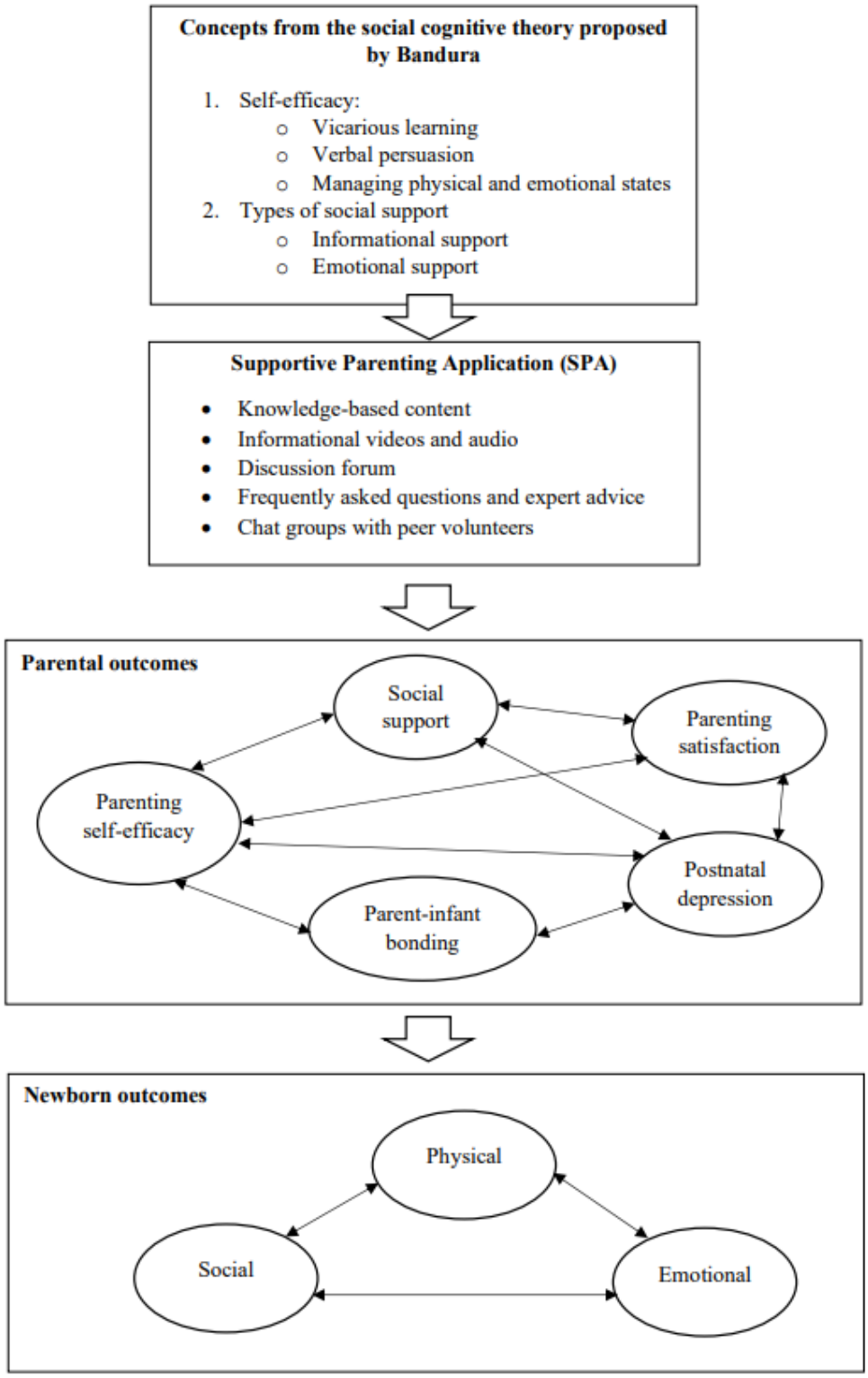
To meet the goals of the rigor cycle, well-established theoretical frameworks were incorporated into the development process of the SPA. The research team referred to the mHealth user engagement pyramid proposed by Singh et al [41] to identify features that could be incorporated into the SPA. Providing educational information is the most fundamental way to engage a patient, and the level of engagement increases with add-on features such as reminders and notifications, records of health information, display and summary of health information, availability of individualized guidance, means of communication between clinicians and family members or caregivers, provision

of support through social networks, and incentivization of behavioral changes [41].

Other theoretical frameworks that guided the development of the SPA's features were the social cognitive theory proposed by Bandura [42,43] and the attachment theory proposed by Bowlby [44]. The social cognitive theory posits that an individual's knowledge acquisition depends on the interaction among the environment, cognitive behavior, and personal factors [43]. An important component of the social cognitive theory is self-efficacy, which is defined as one's belief in one's capability to accomplish a task effectively [42]. Therefore, to ensure successful parenting, Bandura [42,45] emphasized that parents need to possess the self-confidence to perform specific skills and believe that they can achieve their desired outcome. As evidence has shown low levels of parenting self-efficacy as a predictor of postpartum depression [46], the concepts of the self-efficacy theory proposed by Bandura [42,45] were incorporated into the development of the SPA. Mastery of experiences, vicarious experiences, verbal persuasion, and improving physical and emotional states are ways in which self-efficacy can be attained [42]. In addition, Bandura [43] emphasized the importance of social support in achieving self-regulation to aid learning. There are different types of social support (emotional, instrumental, informational, and appraisal) [43]. As the SPA is a mobile app intervention, providing instrumental support was not possible; hence, it focused primarily on emotional and informational support. Appraisal support, which involves the provision of feedback regarding one's performance or qualities, was also not feasible in the context of the SPA. However, specific feedback requested by parents was provided for their individual needs.

According to the attachment theory proposed by Bowlby [44], parenting self-efficacy, social support, and parental emotional well-being are crucial for early parent-infant bonding and infant attachment styles, and these serve as a foundation for the development of positive social relationships in infants. Bowlby [44] postulated that a healthy parent-infant bond may lead to secure attachment, whereas insecurely attached infants may take their insecure patterns (avoidant, anxious ambivalent, or disorganized) with them into adulthood. Congruent to the attachment theory proposed by Bowlby [44], Bandura [45] postulated that self-efficacy will consequently influence parent-infant bonding, which is an important determinant of parental emotional well-being and infant development. Therefore, it is necessary to examine the relationship among parenting self-efficacy, parental outcomes, and infant outcomes. The theoretical framework for this project is presented in Figure 2 [42,43].

Figure 2. Theoretical framework of the Supportive Parenting App [42,43].



Design Cycle: Development Phase

Overview

The research team involved in the development of the SPA comprised maternity health care providers such as obstetricians, midwives, and nurses from the tertiary public hospitals; a psychologist; a pediatrician; and app developers. The intervention consisted of 2 main components: (1) a cloud-based admin portal, which was fully accessible by system admins (midwives and physicians only had access to the frequently asked questions [FAQ] management page) and (2) the SPA, which was accessible by system admins, mothers, fathers, peer volunteers, midwives, and physicians.

The cloud-based admin portal was used to moderate the content in the FAQ section and chat groups, manage user accounts and surveys, and display simple user analytics. The admin portal was developed using Microsoft Visual Studio 2017. The main programming languages used for the back end were ASP.NET

and C#, running on .NET Framework (version 4.5 and above; Microsoft Corporation). The front-end languages used were HTML5, JavaScript, CSS, and jQuery. AJAX techniques were used across the system to facilitate asynchronous data transfer between client and server. The database for the system was Microsoft SQL Server 2016 or later. The system was hosted on Windows Server 2016 or above and Internet Information Services (version 10.0 or above).

The SPA ran on both iOS and Android built using hybrid development tools (Apache Cordova and Ionic Frameworks). All accounts were password protected, and users had to authenticate with their log-in credentials created by the system admin to use the app. Participants were assigned anonymous identifiers, and they remained anonymous throughout.

After the health care research team conveyed the expectations, aims, and thorough overview of the SPA project to the app developers, a logo for the SPA was generated. A pastel-colored theme was chosen to create a soothing and gentle vibe, and the logo displayed an infant lying in the nurturing palm of a parent. The app's aim to highlight the protective and nurturing nature of parents was also demonstrated by the green sprout emerging from the parent's hand that sheltered the infant. The design was finalized after multiple intense discussions between the app developers and the research team; the logo is shown in Figure 3.

The main features of the SPA included knowledge-based content, informational videos and audio clips, a discussion forum, group and private chats with peer volunteers, expert advice from a maternity unit nurse or midwife, and individualized push notifications. Details of the development of each component are described in the following sections.

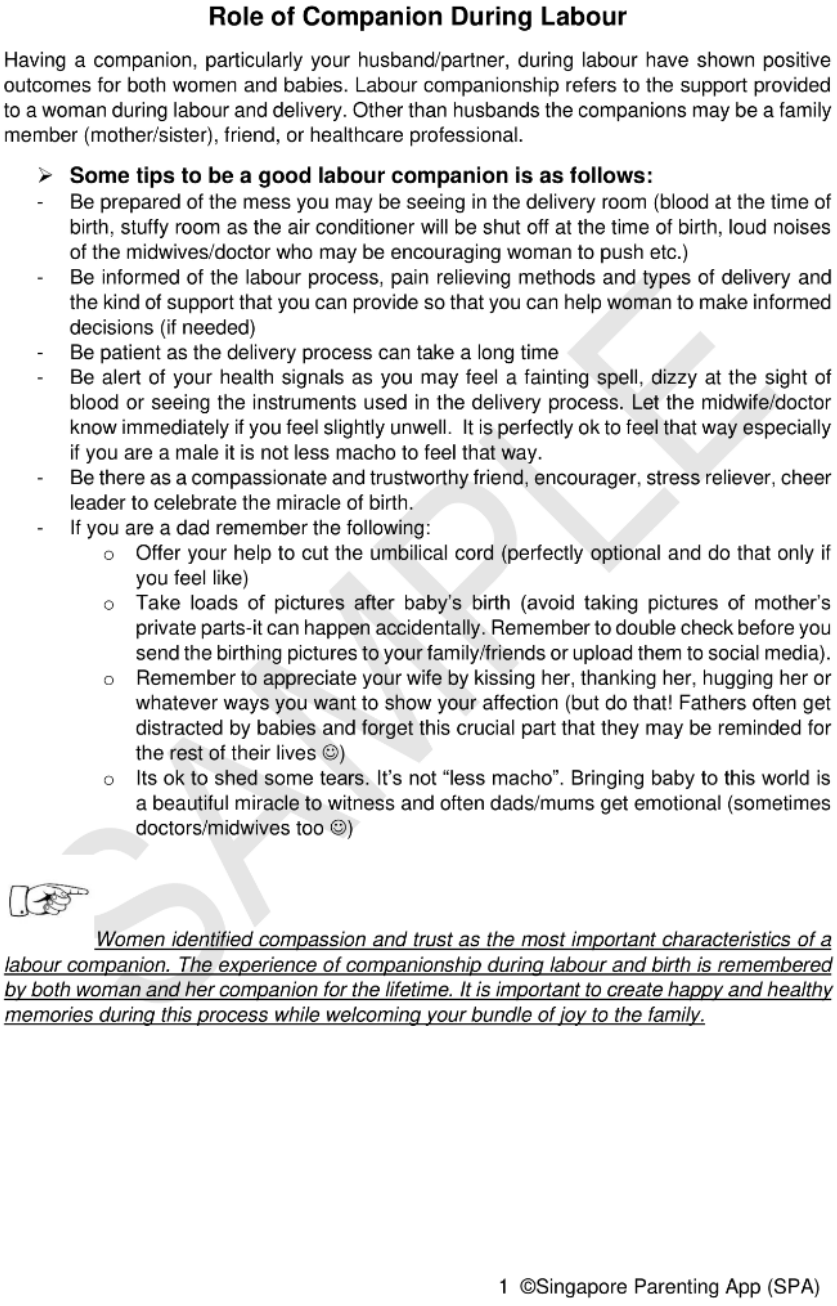
Figure 3. Logo of the Supportive Parenting App.

Knowledge-Based Content

According to Bandura [42,43], the provision of informational support is useful for problem solving and knowledge acquisition. To provide comprehensive information to allay parents' concerns in the perinatal period, knowledge-based content was first brainstormed and then finalized after multiple rounds of intense discussions among a team of health care experts from the psychology, pediatrics, and maternity health care units, as well as a few parent couples. The discussions were held through face-to-face and web-based meetings.

The finalized topics were categorized into 6 groups (general, pregnancy, childbirth, baby care, maternal care, family, and parenthood) to enable users to obtain relevant information. Keyword search was enabled, and the content was sorted by most viewed or most rated. Parents were also allowed to rate the usefulness of the content on a scale of 1 to 5. Knowledge-based content included topics on the prevalence of postpartum depression among mothers and fathers, signs and symptoms of postpartum depression, when to seek help, diet and exercise during pregnancy and after childbirth, preparation for childbirth, types of pain relief during childbirth, newborn and postpartum maternal care, major milestones for the newborn, managing family dynamics, and role of fathers and other caregivers (eg, grandparents) during the perinatal period. There were a total of 42 knowledge-based articles; each knowledge-based article was approximately 500 words long, and the content was written in bullet points to facilitate reading. The *hand symbol* was used consistently to signpost additional parenting tips or highlight instances when parents need to seek emergency help. The English language used was simplified to that of the reading ability of students in primary 6, and this was verified by the publications support unit at the National University of Singapore. An example of a knowledge-based article is shown in Figure 4.

Figure 4. Sample of a knowledge-based article.



Role of Companion During Labour

Having a companion, particularly your husband/partner, during labour have shown positive outcomes for both women and babies. Labour companionship refers to the support provided to a woman during labour and delivery. Other than husbands the companions may be a family member (mother/sister), friend, or healthcare professional.

➢ **Some tips to be a good labour companion is as follows:**

- Be prepared of the mess you may be seeing in the delivery room (blood at the time of birth, stuffy room as the air conditioner will be shut off at the time of birth, loud noises of the midwives/doctor who may be encouraging woman to push etc.)
- Be informed of the labour process, pain relieving methods and types of delivery and the kind of support that you can provide so that you can help woman to make informed decisions (if needed)
- Be patient as the delivery process can take a long time
- Be alert of your health signals as you may feel a fainting spell, dizzy at the sight of blood or seeing the instruments used in the delivery process. Let the midwife/doctor know immediately if you feel slightly unwell. It is perfectly ok to feel that way especially if you are a male it is not less macho to feel that way.
- Be there as a compassionate and trustworthy friend, encourager, stress reliever, cheer leader to celebrate the miracle of birth.
- If you are a dad remember the following:
    - Offer your help to cut the umbilical cord (perfectly optional and do that only if you feel like)
    - Take loads of pictures after baby's birth (avoid taking pictures of mother's private parts-it can happen accidentally. Remember to double check before you send the birthing pictures to your family/friends or upload them to social media).
    - Remember to appreciate your wife by kissing her, thanking her, hugging her or whatever ways you want to show your affection (but do that! Fathers often get distracted by babies and forget this crucial part that they may be reminded for the rest of their lives ☺)
    - Its ok to shed some tears. It's not "less macho". Bringing baby to this world is a beautiful miracle to witness and often dads/mums get emotional (sometimes doctors/midwives too ☺)

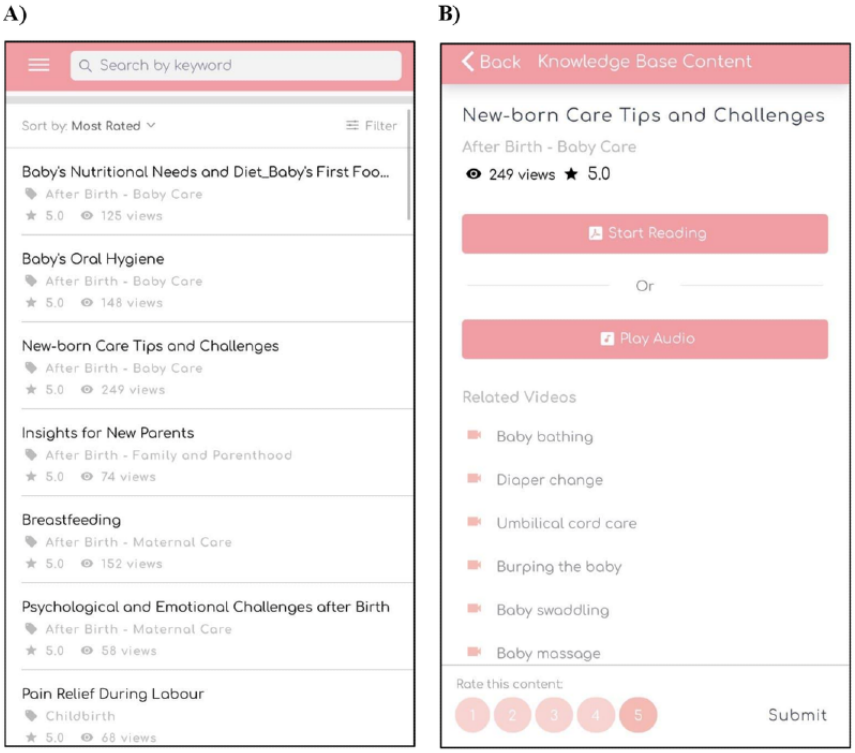
Women identified compassion and trust as the most important characteristics of a labour companion. The experience of companionship during labour and birth is remembered by both woman and her companion for the lifetime. It is important to create happy and healthy memories during this process while welcoming your bundle of joy to the family.

1 ©Singapore Parenting App (SPA)

Informational Videos and Audio Clips

Parenting self-efficacy can be enhanced through vicarious learning or modeling by watching others or personally acquiring a certain behavior successfully [42]. Hence, 27 inspirational and demonstration videos were included in the knowledge-based content section. Video content and scripts were developed by health care experts from different fields. For instance, baby care video scripts were developed by health care professionals from the maternity unit and video scripts on insights into postpartum depression were prepared by psychology experts. The videos consisted of animations, interviews, and skills demonstrations of topics such as personal accounts of mothers who developed and recovered from postpartum depression, diet and exercise during pregnancy, preparation for childbirth, and demonstration of infant care skills and interviews related to a father's role during the perinatal period. Audio versions were also provided to cater to parents who preferred to listen to audio clips or were prohibited from reading or watching videos because of confinement practices. A screenshot of the knowledge-based content page, including videos and audio clips, is shown in Figure 5.

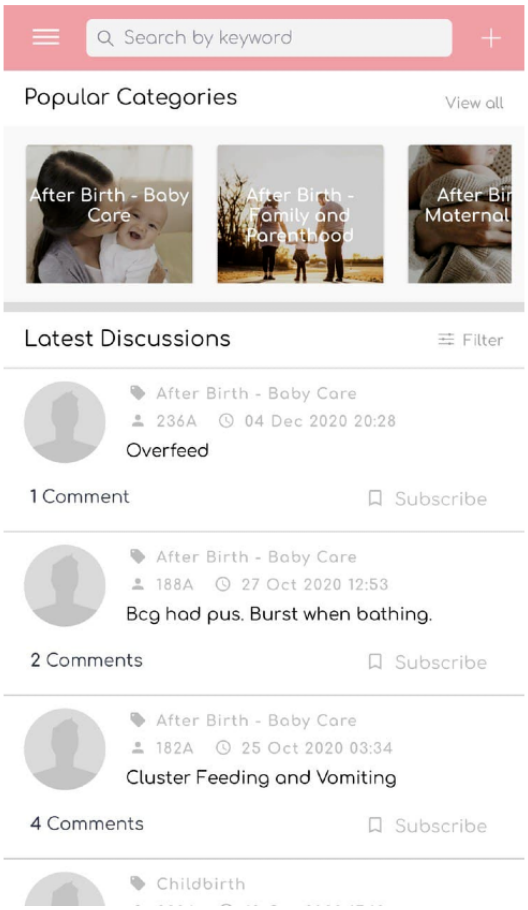
Figure 5. (A) Knowledge-based content page with (B) video and audio materials.



Discussion Forum

Apart from video modeling, a discussion forum was developed to allow parents who faced similar problems to share their tips and experiences. Parents were able to post personal queries and respond to other parents' queries. A midwife or maternity unit nurse was able to moderate the comments to ensure the reliability and accuracy of the information shared. A sample of the discussion forum is shown in Figure 6.

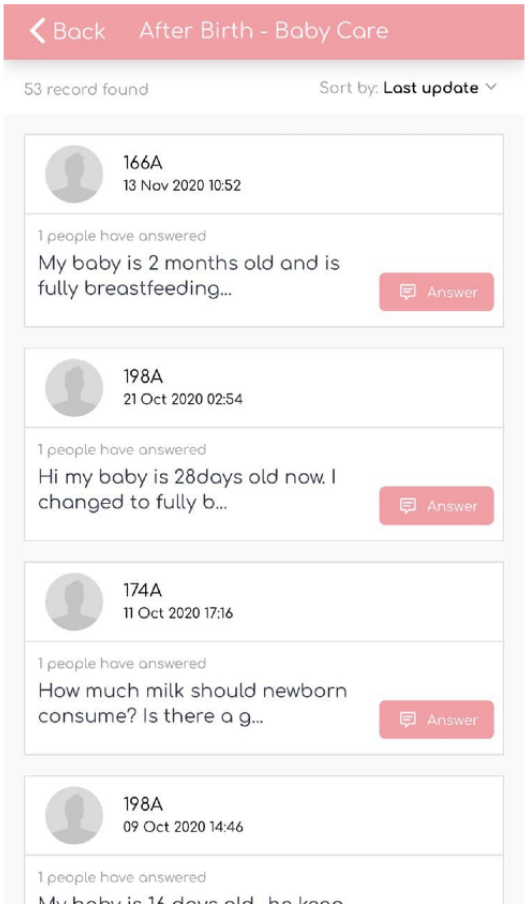
Figure 6. Discussion forum in the Supportive Parenting App.



FAQ and Expert Advice

Further informational support by health care professionals was provided through an FAQ section. If the existing information was insufficient to answer the parents' queries, parents were able to post their questions and upload supporting photos. A screenshot of the FAQ and expert advice section is shown in Figure 7. This feature allowed individualized professional help without parents having to make unnecessary visits to the physician. Expert advice from the maternity unit nurse or midwife was available from pregnancy until 1 month after childbirth. Because of the busy work schedules of the maternity unit nurse and midwife, replies were posted once a day.

Figure 7. Frequently asked questions and experts' advice page.



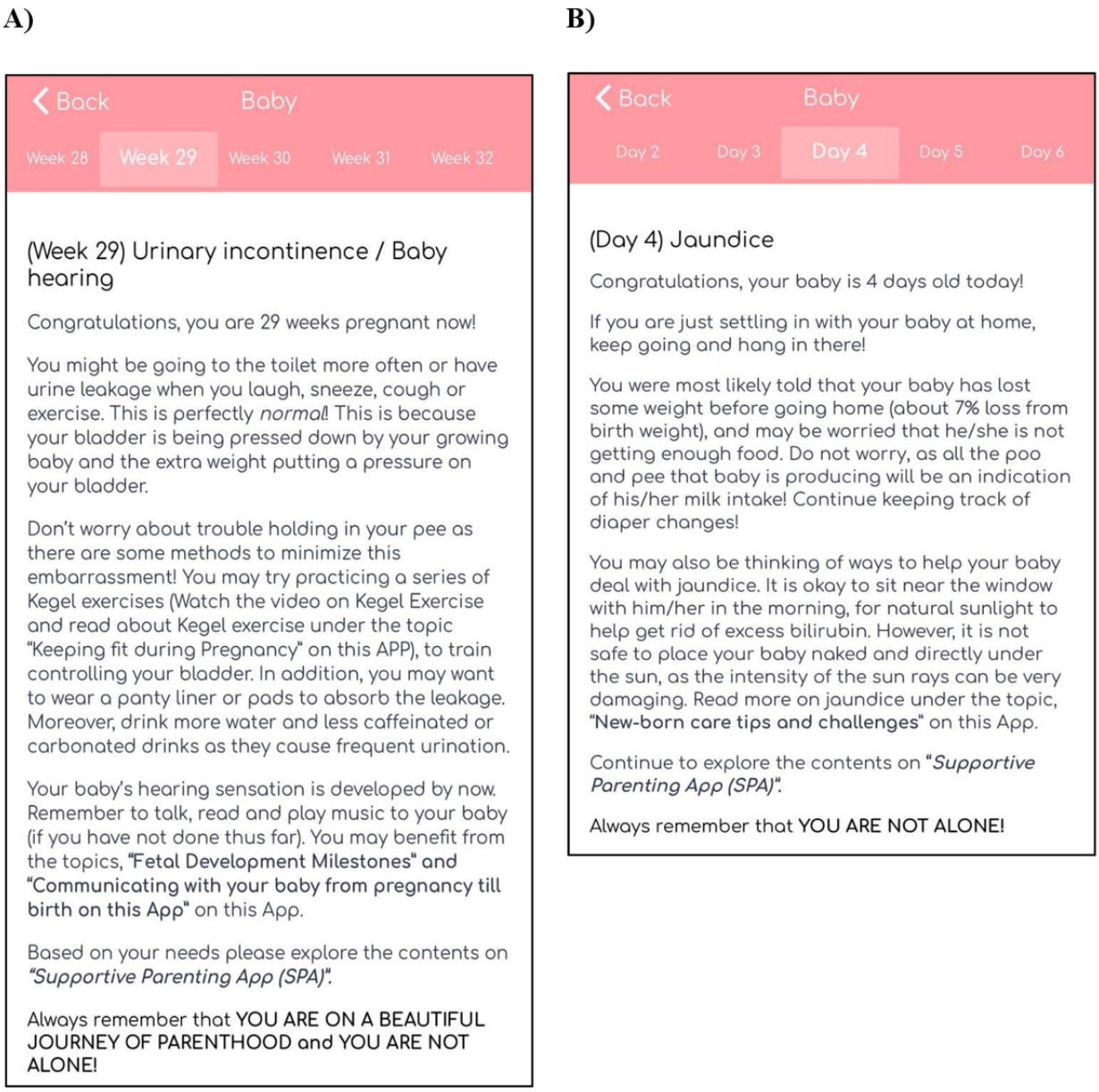
Chat Groups With Peer Volunteers

Mothers were matched for variables not limited to ethnicity and number of children and then assigned to peer volunteers. They were able to seek informational and emotional support from peer volunteers and other mothers or have a private chat with their assigned peer volunteer in the chat groups. Peer volunteers were experienced mothers who had developed and recovered from postpartum depression previously. These peer volunteers were specially trained by a psychiatrist on how to provide adequate support and be a listening ear to needy mothers, as well as on how to navigate the SPA. To ensure fidelity, an electronic training manual was provided to all peer volunteers before the start of a 3-hour web-based training session, and skills acquisition was assessed by the psychologist through web-based role-playing. After the training, each peer volunteer was assigned to approximately five couples (fathers and mothers) who were added to a group chat. Peer volunteers were encouraged to check up on the participants regularly through the group chat and to encourage mothers to interact and share their experiences. Parents who were uncomfortable with expressing themselves in a group setting had the choice to send private messages to their peer volunteer. Overall, the role of the peer volunteer was to provide emotional support to needy mothers, help foster peer support, and develop a close-knit community among parents in the chat group.

Push Notifications

Finally, push notifications were sent on a weekly basis during pregnancy, daily during the first month of the postpartum period, and biweekly thereafter until 6 months after childbirth. These notifications provided individualized information to parents, according to the specific week of their pregnancy and the age of their newborn after childbirth. The push notifications during pregnancy helped to track the pregnancy stages and provided information specific to the mother's pregnancy stage, such as common physical experiences (eg, Braxton Hicks contraction), ways to manage physical symptoms, relevant informational material that parents could refer to in the SPA, and words of encouragement. Push notifications that were sent after childbirth kept both parents updated on the infant's developmental milestones, informed parents on common physical occurrences and changes in the mother (eg, vaginal bleeding) and the infant (eg, weight loss), presented ways to care for both mother and infant, provided supportive tips for fathers, guided parents to relevant information in the SPA, and included words of encouragement. Parents were also advised to look out for required information from the knowledge-based content and informational videos and audio clips according to their individualized needs and not to switch off the notifications because this could interfere with the intervention. Examples of push notifications delivered during pregnancy and after childbirth are presented in Figure 8.

Figure 8. Examples of push notifications delivered (A) during pregnancy and (B) after childbirth.



Design Cycle: Evaluation

Pilot Test

To assess the SPA's features, functionality, usability, and content accuracy, a pilot test for the mobile app and administrator webpage was conducted among users (N=10) who were new parents and research team members (ie, app developers, clinicians, and research assistants). Feedback was gathered qualitatively and discussed among team members at a web-based meeting to further refine the app and optimize the user experience. This process was repeated multiple times and ceased when all the research team members arrived at a mutual agreement and were satisfied with the finalized version of the app. The mobile app was then made available to participants in the iOS and Android app stores.

System Compatibility

During the initial piloting, there were issues regarding the viewer compatibility of the cloud-based admin portal on different web browsers. The admin portal was not accessible on Internet Explorer and could only be viewed using Google Chrome, Mozilla Firefox, or Microsoft Edge. This was subsequently made known to all system admin users. Similarly, the SPA was initially incompatible with the older Android OS versions, but updates were made by the developer to support the older OS versions.

Restricted Admin Functions and Data Analytics

In addition, the admin portal did not allow for a full overview of and access to the SPA's contents, especially in terms of group chat details. System admin users could only view peer volunteer group chat conversations using the mobile app. In addition, the initial data analytics function on the admin portal only displayed aggregated data on user frequency. However, individual user

frequency data were deemed more informative for future research purposes. As subsequent requests for collating individual user data by the research team resulted in extra development charges and time, no enhancement was made to the data analytics.

App Notifications

The first testing of the SPA revealed that mothers did not receive group chat notifications. This was quickly fixed by the app developers so that mothers could receive a pop-up notification whenever somebody posted in the group chat. However, the problem persisted in the second round of user testing. Group chat notifications were not received when the app was not constantly updated, and parents did not receive any prompt to update the app. This situation was later rectified by sending reminders to update the app. Similarly, team members in charge of the expert advice section made a request for pop-up notifications to enable them to respond to parents' queries promptly. However, this was not implemented because of the project's tight budget. After the first user testing, educational contents (PDF files and videos) were not linked to the push notifications as expected. This issue was then rectified by the app developers for parents' convenience.

Search Function

The research team also highlighted issues with the search function of the knowledge-based content. Initially, the search engine was only able to detect specific keywords in the title but not in the text. The function for in-text search required the additional purchase of plug-ins and further development of codes that involved time and cost that the team did not take into account initially.

Discussion

Principal Findings

The development process of the SPA, which included requirement identification, the use of theories and past artifacts, the design of artifacts, and user testing, was well documented using the ISR framework. Although the ISR framework was able to capture the design technicalities and rigor of the project, it failed to highlight interpersonal and real-life challenges. Apart from technical and user issues, the design cycle faced prominent issues such as logistical limitations and also in terms of teamwork and communication because of the effects of the COVID-19 outbreak, which affected the process timeline.

Challenges in the Design Cycle: Logistics Limitations

With respect to the FAQ and expert advice section, the research team ideally wanted to provide a readily available hotline or personalized help for parents in need. However, because of limited resources and other work commitments, it was not feasible for the research team to be on constant standby to respond to parents' queries immediately. Hence, a decision was made to have a maternity unit nurse or midwife to check and respond to queries posted in the FAQ section once a day and respond within 24 hours of the queries raised by the parents. Parents with urgent queries were recommended to call the hospital or visit the physician immediately. Despite the team's best efforts to provide parents with a personalized, fast, trusted, and reliable information source, the strain on logistics and manpower led the research team to consider the future incorporation of live chatbots in the mobile app. The use of chatbots or conversational agents in the health care setting has been on the rise, especially with smartphone apps, in the areas of patient treatment and monitoring, health care service support, and patient education [47]. Given the novelty of the use of chatbots, there is a paucity of randomized controlled trials evaluating their efficacy in providing knowledge-based information to parents in the perinatal period. However, promising user test results from a few development and usability studies [48-50] indicate the viability of chatbots, thus paving the way for future research.

Second, the recruitment and training of peer volunteers posed a challenge because peer volunteers were also mothers who had work and family commitments; hence, it was difficult to recruit dedicated peer volunteers to take on this yearlong project. In addition, the conflicting schedules of peer volunteers resulted in a few postponements of face-to-face training sessions until the COVID-19 outbreak led to their conversion to web-based sessions instead. Team members in charge of conducting the sessions had to rapidly adapt to the situation and convert the face-to-face training to a web-based session. Research has shown that the efficacy of web-based interpersonal skills (ie, communication, counseling, and interview) training on skills acquisition is on par with face-to-face training [51-53]. Furthermore, participants have also commended web-based courses for their better accessibility and the ease of remote learning [54,55]; thus, a web-based session may better cater to volunteers who are working mothers. However, Doo [56] has argued that cognitive understanding of interpersonal skills does not translate into successful skills execution and interpersonal skills can only be acquired through real-life practice, which is difficult in web-based learning environments. Although web-based training may not be as beneficial as in-person training in terms of building rapport among peer volunteers and conducting role-plays, this was the optimal resolution in times of a COVID-19 outbreak. We are waiting to interview the peer volunteers at the completion of this project about their perceptions of receiving the web-based training. The interviews may provide true evaluation and recommendations on enhancing such web-based training in the future.

Challenges in the Design Cycle: Teamwork and Communication

As this was intended to be a multisite study, extensive collaboration among health care professionals from various institutions was required. However, a key challenge faced during the Tuckman model's forming phase of team development [38] was the assembling and scheduling of face-to-face meetings with different experts from various health care institutions, given their hectic schedules. As face-to-face meetings were crucial at the initial stage for the research team to get acquainted and share their expertise and viewpoints, meeting dates were scheduled when there was a likelihood of the highest possible attendance. Team members who were unable to attend received the meeting materials and minutes, and any further feedback was communicated through email. After the delegation of the

main roles and responsibilities, subsequent follow-up meetings, updates, and discussions were conducted through web-based meetings with respective parties. At the storming phase, given each team member's field of expertise and the dearth of essential information, it was difficult for the team to agree on the finalized topics and knowledge-based content to be included in the app. However, the team managed to ascend to the norming phase through resolution by constructive negotiation and eventual mutual agreement to include less essential information in the FAQ section than in the knowledge-based content section. During the performing phase, the team managed to work together effectively and cohesively and successfully rolled out the SPA for user acceptability testing. Overall, the Tuckman model of team development highlighted key characteristics and conflict resolution strategies that enabled optimal team development, which future mobile app developmental studies can use.

Another prominent issue concerned the communication between the app developers and the health care research team. Although the research team comprised experts in content development, they were unfamiliar with technical programming terms and were unable to communicate effectively with the app developers on their app preferences such as cloud hosting, data analytics function, and security mechanisms of the app. This communication difficulty resulted in different expectations for the app and led to last-minute enhancements that incurred additional costs that were not initially considered. Therefore, both parties should be careful with the use of technical jargon to avoid confusion and to constantly clarify any uncertainty or details. Alternatively, visual diagrams (eg, mind maps, drawings, and sample screenshots) could be useful in visualizing and conveying expectations of specific mobile app features.

Implications for Future Studies

Much can be learned from the intervention development process of the SPA based on the aforementioned challenges. A major setback was the sudden COVID-19 outbreak, which interrupted the development of the intervention and delayed the project. The flexibility and quick adaptability of the team in finding solutions was crucial in ensuring the continuity of the project. Although face-to-face meetings were impossible, the team was able to conduct meetings through web-based video calls and follow-up emails. Face-to-face training for peer volunteers was also revised and conducted on the web. Team cohesiveness and adaptability were crucial in effectively managing unforeseen circumstances.

Team cohesion is dependent on open and clear communication among team members to ensure that everyone's expectations are aligned, uncertainties are clarified, and there is transparency for group decision-making. This is especially crucial between the research team and the app developers because their areas of expertise differ and miscommunication in expectations could lead to unfavorable outcomes. However, the interdisciplinary, team-based approach ensured the success of the app development in this study. In addition, a structured and effective workflow was implemented through strict planning of timelines, delegation of responsibilities to team members, close updates and timely feedback among team members, and the charting of project milestones using the Gantt chart. The physical visualization of the project timeline kept everyone in check and allowed better handling of delays.

Another critical lesson learned was to ensure sufficient financial and time budgeting for last-minute enhancements of the intervention. Because of project delays during the COVID-19 pandemic, the team could not spend more time and money for potential feature enhancements that may influence the detailed data analytics. As the basic infrastructure for the parenting app has already been established and the team members have acquired sufficient experience, incorporating future improvements such as adding additional features to the platform would require much less time than that needed for the initial project. Future studies developed as part of this SPA project could assess actual app use, cost-effectiveness, impacts on longitudinal parental and infant outcomes, and qualitative user experience. The use of chatbots to answer parental queries on a timely basis could also be incorporated and evaluated in future research on supportive apps for parents.

Conclusions

This study presents a multistage iterative development process of an mHealth SPA using the ISR framework. Rooted in social cognitive theory and attachment theory, the SPA offers parents a holistic means of obtaining informational and emotional support from knowledge-based content, parents with similar backgrounds, trained peer volunteers, and health care professionals. However, the development process had its challenges because the design cycle involved other logistical and team communication issues that were not taken into consideration in the ISR framework. These issues can potentially be resolved with adequate time and financial budgeting and by developing team cohesion.

Acknowledgments

This study was funded by the Ministry of Health (HSRGHP18may-0001). The authors would like to thank the National University Health System Research Support Unit for their assistance in the language editing of this manuscript.

Conflicts of Interest

None declared.

References

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2020 Oct 15;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9]
2. Cox JL, Murray D, Chapman G. A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry 1993 Jul 02;163(1):27-31. [doi: 10.1192/bjp.163.1.27] [Medline: 8353695]
3. Davé S, Petersen I, Sherr L, Nazareth I. Incidence of maternal and paternal depression in primary care: a cohort study using a primary care database. Arch Pediatr Adolesc Med 2010 Nov 01;164(11):1038-1044. [doi: 10.1001/archpediatrics.2010.184] [Medline: 20819960]
4. Doss BD, Rhoades GK, Stanley SM, Markman HJ. The effect of the transition to parenthood on relationship quality: an 8-year prospective study. J Pers Soc Psychol 2009 Mar;96(3):601-609 [FREE Full text] [doi: 10.1037/a0013969] [Medline: 19254107]
5. Ghaedrahmati M, Kazemi A, Kheirabadi G, Ebrahimi A, Bahrami M. Postpartum depression risk factors: a narrative review. J Educ Health Promot 2017;6:60 [FREE Full text] [doi: 10.4103/jehp.jehp_9_16] [Medline: 28852652]
6. Yim IS, Stapleton LR, Guardino CM, Hahn-Holbrook J, Schetter CD. Biological and psychosocial predictors of postpartum depression: systematic review and call for integration. Annu Rev Clin Psychol 2015;11:99-137 [FREE Full text] [doi: 10.1146/annurev-clinpsy-101414-020426] [Medline: 25822344]
7. Goodman JH. Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. J Adv Nurs 2004 Jan;45(1):26-35. [doi: 10.1046/j.1365-2648.2003.02857.x] [Medline: 14675298]
8. Fisher SD. Paternal mental health: why is it relevant? Am J Lifestyle Med 2017;11(3):200-211 [FREE Full text] [doi: 10.1177/1559827616629895] [Medline: 30202331]
9. Paulson JF, Bazemore SD. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. J Am Med Assoc 2010 May 19;303(19):1961-1969. [doi: 10.1001/jama.2010.605] [Medline: 20483973]
10. Letourneau NL, Dennis C, Benzies K, Duffett-Leger L, Stewart M, Tryphonopoulos PD, et al. Postpartum depression is a family affair: addressing the impact on mothers, fathers, and children. Issues Ment Health Nurs 2012 Jul;33(7):445-457. [doi: 10.3109/01612840.2012.673054] [Medline: 22757597]
11. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet 2014 Nov;384(9956):1800-1819. [doi: 10.1016/S0140-6736(14)61277-0]
12. Goodman SH, Gotlib IH. Risk for psychopathology in the children of depressed mothers: a developmental model for understanding mechanisms of transmission. Psychol Rev 1999 Jul;106(3):458-490. [doi: 10.1037/0033-295x.106.3.458] [Medline: 10467895]
13. Luoma I, Tamminen T, Kaukonen P, Laippala P, Puura K, Salmelin R, et al. Longitudinal study of maternal depressive symptoms and child well-being. J Am Acad Child Adolesc Psychiatry 2001 Dec;40(12):1367-1374. [doi: 10.1097/00004583-200112000-00006] [Medline: 11765281]
14. Rifkin-Graboi A, Bai J, Chen H, Hameed WB, Sim LW, Tint MT, et al. Prenatal maternal depression associates with microstructure of right amygdala in neonates at birth. Biol Psychiatry 2013 Dec 01;74(11):837-844. [doi: 10.1016/j.biopsych.2013.06.019] [Medline: 23968960]
15. Kane P, Garber J. The relations among depression in fathers, children's psychopathology, and father-child conflict: a meta-analysis. Clin Psychol Rev 2004 Jul;24(3):339-360. [doi: 10.1016/j.cpr.2004.03.004] [Medline: 15245835]
16. Dennis C, Kingston D. A systematic review of telephone support for women during pregnancy and the early postpartum period. J Obstet Gynecol Neonatal Nurs 2008 May;37(3):301-314. [doi: 10.1111/j.1552-6909.2008.00235.x] [Medline: 18507601]
17. Dennis C, Creedy D. Psychosocial and psychological interventions for preventing postpartum depression. Cochrane Database Syst Rev 2004 Oct 18;4(4):CD001134. [doi: 10.1002/14651858.CD001134.pub2] [Medline: 15495008]
18. Van Zeijl J, Mesman J, Van IJzendoorn MH, Bakermans-Kranenburg MJ, Juffer F, Stolk MN, et al. Attachment-based intervention for enhancing sensitive discipline in mothers of 1- to 3-year-old children at risk for externalizing behavior problems: a randomized controlled trial. J Consult Clin Psychol 2006;74(6):994-1005. [doi: 10.1037/0022-006x.74.6.994]
19. Lavender T, Walkinshaw SA. Can midwives reduce postpartum psychological morbidity? A randomized trial. Birth 1998 Dec 04;25(4):215-219. [doi: 10.1046/j.1523-536x.1998.00215.x] [Medline: 9892887]
20. Shorey S, Chan SW, Chong YS, He H. A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. J Adv Nurs 2015 Jun 15;71(6):1260-1273. [doi: 10.1111/jan.12590] [Medline: 25496615]
21. Small R, Lumley J, Donohue L, Potter A, Waldenström U. Randomised controlled trial of midwife led debriefing to reduce maternal depression after operative childbirth. Br Med J 2000 Oct 28;321(7268):1043-1047 [FREE Full text] [doi: 10.1136/bmj.321.7268.1043] [Medline: 11053173]
22. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The empirical evidence for telemedicine interventions in mental disorders. Telemed J E Health 2015 Dec 01:87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]
23. Tehrani N. How advances in digital health benefit pregnant women. Int J Emerg Res Manag Technol 2017 Apr 20;6(4):8-10. [doi: 10.23956/ijermt/sv6n4/185]

24. Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]
25. Naser E, Mackey S, Arthur D, Klainin-Yobas P, Chen H, Creedy DK. An exploratory study of traditional birthing practices of Chinese, Malay and Indian women in Singapore. Midwifery 2012 Dec;28(6):865-871. [doi: 10.1016/j.midw.2011.10.003] [Medline: 22079015]
26. Ong SF, Chan WS, Shorey S, Chong YS, Klainin-Yobas P, He H. Postnatal experiences and support needs of first-time mothers in Singapore: a descriptive qualitative study. Midwifery 2014 Jun;30(6):772-778. [doi: 10.1016/j.midw.2013.09.004] [Medline: 24161493]
27. Danbjørg DB, Wagner L, Kristensen B, Clemensen J. Intervention among new parents followed up by an interview study exploring their experiences of telemedicine after early postnatal discharge. Midwifery 2015 Jun;31(6):574-581. [doi: 10.1016/j.midw.2015.02.007] [Medline: 25765743]
28. Leroux E, Rivas H. Mobile health without borders (NovoEd). Stanford University. 2013. URL: https://www.mooc-list.com/course/mobile-health-without-borders-novoed [accessed 2021-12-10]
29. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
30. Virani A, Duffett-Leger L, Letourneau N. Parenting apps review: in search of good quality apps. Mhealth 2019;5:44 [FREE Full text] [doi: 10.21037/mhealth.2019.08.10] [Medline: 31620471]
31. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019 Jun 17;2(1):54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
32. Annual survey on infocomm usage in households and by individuals for 2019. Infocomm and Media Development Authority. 2019. URL: https://www.imda.gov.sg/-/media/Imda/Files/Infocomm-Media-Landscape/Research-and-Statistics/Survey-Report/2019-HH-Public-Report_09032020.pdf [accessed 2021-12-10]
33. Smartphone market in Singapore - statistics and facts. Statista. 2020. URL: https://www.statista.com/topics/5842/smartphones-in-singapore/ [accessed 2021-12-10]
34. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 08;10(4):e033516 [FREE Full text] [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
35. Hevner A. A Three Cycle View of Design Science Research. Scandinavian J Inform Syst 2007;19(2):1-7 [FREE Full text]
36. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
37. Eaton LH, Doorenbos AZ, Schmitz KL, Carpenter KM, McGregor BA. Establishing treatment fidelity in a web-based behavioral intervention study. Nurs Res 2011;60(6):430-435 [FREE Full text] [doi: 10.1097/NNR.0b013e31823386aa] [Medline: 22048559]
38. Tuckman BW. Developmental sequence in small groups. Psychol Bull 1965 Jun;63(6):384-399. [doi: 10.1037/h0022100] [Medline: 14314073]
39. Shorey S, Ng ED. Evaluation of a technology-based peer-support intervention program for preventing postnatal depression (Part 2): qualitative study. J Med Internet Res 2019 Aug 29;21(8):e12915 [FREE Full text] [doi: 10.2196/12915] [Medline: 31469080]
40. Shorey S, Yang YY, Dennis C. A mobile health app-based postnatal educational program (Home-but not Alone): descriptive qualitative study. J Med Internet Res 2018 Apr 19;20(4):e119 [FREE Full text] [doi: 10.2196/jmir.9188] [Medline: 29674314]
41. Singh K, Drouin K, Newmark L, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue brief (The Commonwealth Fund). 2016. URL: https://www.issuelab.org/resources/25066/25066.pdf [accessed 2021-12-10]
42. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
43. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. New Jersey: Englewood Cliffs; 1985.
44. Bowlby J. Attachment and Loss. London: Hogarth P; 1969.
45. Bandura A. Self-Efficacy: The Exercise of Control. New York: W. H. Freeman; 1997.
46. Shorey S, Ng YP, Ng ED, Siew AL, Mörelius E, Yoong J, et al. Effectiveness of a technology-based supportive educational parenting program on parental outcomes (Part 1): randomized controlled trial. J Med Internet Res 2019 Feb 13;21(2):e10816 [FREE Full text] [doi: 10.2196/10816] [Medline: 30758289]
47. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]

48. Chung K, Cho HY, Park JY. A chatbot for perinatal women's and partners' obstetric and mental health care: development and usability evaluation study. JMIR Med Inform 2021 Mar 03;9(3):e18607 [FREE Full text] [doi: 10.2196/18607] [Medline: 33656442]
49. Sadavarte S, Bodanese E. Pregnancy companion chatbot using Alexa and Amazon web services. In: Proceedings of the IEEE Pune Section International Conference (PuneCon). 2019 Presented at: IEEE Pune Section International Conference (PuneCon); Dec. 18-20, 2019; Pune, India. [doi: 10.1109/punecon46936.2019.9105762]
50. Vaira L, Bochicchio M, Conte M, Casaluci F, Melpignano A. MamaBot: a system based on ML and NLP for supporting women and families during pregnancy. In: Proceedings of the 22nd International Database Engineering & Applications Symposium. 2018 Presented at: IDEAS 2018: 22nd International Database Engineering & Applications Symposium; June 18 - 20, 2018; Villa San Giovanni, Italy p. 273-277. [doi: 10.1145/3216122.3216173]
51. Kyaw BM, Posadzki P, Paddock S, Car J, Campbell J, Tudor Car L. Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Aug 27;21(8):e12967 [FREE Full text] [doi: 10.2196/12967] [Medline: 31456579]
52. Neuhauser C. Learning style and effectiveness of online and face-to-face instruction. Am J Dist Edu 2002 Jun;16(2):99-113. [doi: 10.1207/s15389286ajde1602_4]
53. Murdock J, Williams A, Becker K, Bruce M, Young S. Online versus on-campus: a comparison study of counseling skills courses. J Hum Resour Adult Learn 2012;8(1):1-14 [FREE Full text]
54. Mukhtar K, Javed K, Arooj M, Sethi A. Advantages, limitations and recommendations for online learning during COVID-19 pandemic era. Pak J Med Sci 2020 May 18;36(COVID19-S4). [doi: 10.12669/pjms.36.covid19-s4.2785]
55. Appanna S. A review of benefits and limitations of online learning in the context of the student, the instructor and the tenured faculty. Int J E-Learn 2008;7(1):5-22 [FREE Full text]
56. Doo MY. A problem in online interpersonal skills training: do learners practice skills? Open Learn: J Open Dist e-Learn 2007 Jan 23;21(3):263-272. [doi: 10.1080/02680510600953252]

Abbreviations

FAQ: frequently asked questions
ISR: information systems research
mHealth: mobile health
SPA: Supportive Parenting App

Edited by R Kukafka; submitted 08.01.21; peer-reviewed by R Otte, B Chaudhry; comments to author 19.02.21; revised version received 05.03.21; accepted 10.11.21; published 24.12.21.

Please cite as:
Shorey S, Tan TC, thilagamangai, Mathews J, Yu CY, Lim SH, Shi L, Ng ED, Chan YH, Law E, Chee C, Chong YS
Development of a Supportive Parenting App to Improve Parent and Infant Outcomes in the Perinatal Period: Development Study
J Med Internet Res 2021;23(12):e27033
URL: https://www.jmir.org/2021/12/e27033
doi:10.2196/27033
PMID:36260376

Original Paper

Expectations of Health Researchers From Academic Social Network Sites: Qualitative Study

Mohammad Dehghani[1], PhD; Mehdi Kahouei[2], PhD; Shahin Akhondzadeh[3], PhD; Bita Mesgarpour[4], PhD; Reza Ferdousi[5], PhD

[1]Department of Health Information Technology, Khomein University of Medical Sciences, Khomein, Iran

[2]Social Determinants of Health Research Center, School of Allied Medical Sciences, Semnan University of Medical Sciences, Semnan, Iran

[3]Psychiatric Research Center, Roozbeh Hospital, Tehran University of Medical Sciences, Tehran, Iran

[4]National Institute for Medical Research Development (NIMAD), Tehran, Iran

[5]Department of Health Information Management, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran

Corresponding Author:
Reza Ferdousi, PhD
Department of Health Information Management
School of Management and Medical Informatics
Tabriz University of Medical Sciences
Daneshgah St
Tabriz, 5165665811
Iran
Phone: 98 9144024707
Fax: 98 4133251378
Email: ferdousi.r@gmail.com

Abstract

Background: Today, academic social network sites' role in improving the quality of education and how investigators conduct their research has become more critical.

Objective: This study aimed to investigate Iranian health researchers' requirements for academic social network sites from a low-income country perspective.

Methods: This qualitative study with a phenomenological approach was done in 2020. In this study, 23 researchers in the health system were selected by purposive sampling. Semistructured interviews were used to collect data. Data were analyzed by MaxQDA-10 software and the content analysis method.

Results: We identified 2 categories of functional and technical characteristics in the study participants' expectations. Functional characteristics included facilitating communication and team activities, managing scientific publications, enhancing the process of conducting research, being informative, and sharing and trading laboratory materials and equipment. Technical characteristics of an academic social network include user management capabilities, high security and privacy, being user-friendly, and other technical features.

Conclusions: Health researchers emphasized 2 functional and technical characteristics required to meet academic social network sites' expectations.

(J Med Internet Res 2021;23(12):e24643) doi:10.2196/24643

KEYWORDS

research; social network; academic social network; research network; academic; researcher; literature; qualitative; content analysis

Introduction

Nonacademic social networking sites such as Facebook are prevalent, and researchers can use them. However, studies show challenges and restrictions for academic users on these sites [1-4]. Today, academic social network sites (ASNSs) have become an integral part of researcher work [5,6]. An ASNS is a type of internet service that facilitates communication between researchers [7], shares scientific resources (news, reports, articles, and data sets), exchanges research opinions, and informs about the current research trend [8]. In addition to publishing researchers' work and facilitating personal exchanges, ASNSs

are tools for describing organizational information and researcher interests [9].

Ijad Madisch, one of the creators of Research Gate, acknowledges that ASNS promotes transparency in the research process and ultimately leads to the strengthening of scientific research [10]. In May 2017, Alexa.com ranked globally Research Gate and Academia.edu 321st and 577th, respectively, indicating increased use of ASNSs [11]. Based on Dong [12], ASNSs have a positive impact on the performance of academics. However, Salvation's study [13] in Malaysia found hidden weaknesses of ASNS.

Every ASNSs is customized for one or more specific purposes; for example, Research Gate is primarily for contacting colleagues and counseling. Mendeley offers the opportunity to receive new articles [14].

Along with researchers from other countries, Iranian health researchers use different ASNSs to conduct their research activities [15]. However, 61% of Iranian researchers do not trust this social network [16]. Ghazimirsaeed [17] examined the use of the academic social networks in Iran and showed 83% (44/53) of Iranian medical science universities were present in the ASNS in 2017. On average, 180 researchers from each university and 1161 departments of the medical universities were members of these ASNSs [17].

ASNSs are created professionally and with a specific purpose, and each of them has its particular users [18]. Investigating the needs and expects of researchers from ASNSs can increase the use of these social networks and make them successful. This study aimed to investigate health researchers' requirements from ASNS in Iran, as a low-income country. In addition to being used in the design of ASNS, this study's results can strengthen them.

Methods

Study Design

This qualitative study to identify Iranian researchers' expectations from ASNS was done in 2020. This article is excerpted from a doctoral dissertation entitled “Designing and Implementing a Social Network for Laboratory Researchers in Health” [19]. We chose the qualitative method to highlight participants' experiences, knowledge, and silent information [20,21]. We selected the phenomenological approach due to the lack of a complete theory of the expectations of Iranian researchers from ASNS [22].

Participants and Setting

The study environment was a research center affiliated with medical colleges. Research managers, faculty members, postdoctoral researchers, and PhD students participated in this study. The inclusion criteria were the membership in 2 or more academic social networks and updating their user profile on academic social networks at least once a month. Participants were selected by the purposive sampling method. Individuals with good information and who provided their information appropriately were chosen as participants in this method [23-25].

The interviews with participants continued until information saturation was achieved, and researchers felt that new information on new participants was not available; this step was achieved with 23 participants. After the study's initiation and interviews, we used theoretical sampling to identify people who could provide rich and beneficial information for researchers.

We attempted to have a diversity of age, employment status, work experience, degree, and job position in our study sample.

Data Collection

Semistructured interviews in the Persian language have been done conveniently for participants. The interview questions were developed by using literature reviews and expert opinions and comprised of five questions. The interviewee was initially asked to introduce themselves and explain their recent research activities. In the second question, the researcher was asked which academic social networks they use and why; and which features are interesting to them.

The next question was about the advantages and disadvantages of these social networks. In that question, participants were asked to compare 2 or more academic social networks they have used. The fourth question asked how academic social networks could accelerate the research process and improve their quality, and the final question queried which features would be considered if the researcher were to design an ASNS.

We used S-recorder software (version 20.1.186.12; Samsung) to record the interviews alongside note-taking. Each interview lasted between 30 and 42 minutes.

Qualitative Analysis

The respondents' answers were immediately typed, summarized, and reviewed several times by listening to and reading the primary information.

Conventional content-method and MaxQDA-10 software (version 10; VERBI) were used for data analysis. The conventional content method is very useful for identifying, analyzing, and reporting the patterns (themes) in qualitative studies [26-28]. Respondent validity and immersed expert and peer checks were used for data portability, rigor, and reliability.

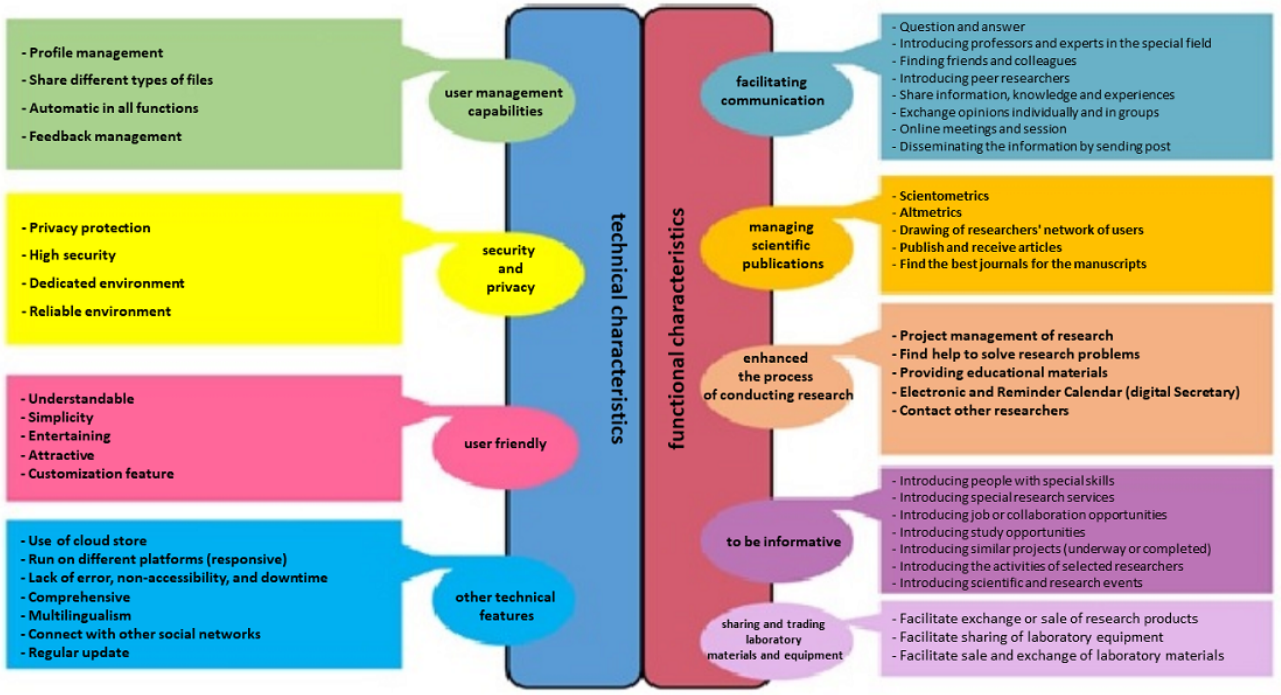
Following informed consent procedures, the research participants are provided with a brief verbal explanation of the study and told that they could leave the study at any time. The participant code and their job were used to report their statements to keep the information confidential. The ethics committee at Tabriz University of Medical Sciences approved this study (IR.TBZEDMED.REC.1398.184).

Results

Overview

Study participants included 7 research managers, 6 faculty members, 4 postdoctoral researchers, and 6 PhD students. Based on the interviews, the researchers' expectations from the ASNSs were divided into 2 general categories: the system's functional and technical characteristics (Figure 1).

Figure 1. Health researchers' expectations for academic social network sites.



Functional Characteristics

Functional characteristics included facilitating communication and team activities, managing scientific publications, enhancing the process of conducting research, being informative, and sharing and trading laboratory materials and equipment. Functional characteristics have 5 main themes and 29 subthemes. Based on the number of repetitions of the subtheme in the interviews, health researchers' most important expectation from ASNSs was to contact other researchers. Finding help to solve research problems, contacting other researchers, and introducing the activities of selected researchers are essential ASNS expectations for PhD students (Table 1).

According to the participants' scientific rank, the four columns of Tables 1 and 2 specify how many people have mentioned the relevant theme in their speeches. Moreover, the total number of people who have mentioned a particular theme is specified in the last column of the tables.

Table 1. Functional expectations of health researchers from academic social network sites.

Theme and subthemes	Description	PhD student (n=6)	Postdoctoral researcher (n=4)	Faculty member (n=6)	Research manager (n=7)	All (n=23)
Facilitating communication and team activities						
Question and answer	Health researchers need to receive advice from experts through ASNS to solve their problems.	3	2	3	1	9
Introducing professors and experts in the special field	Health researchers would like to get acquainted with the best professors and experts in their research field using ASNS.	2	1	4	0	8
Finding friends and colleagues	ASNS must allow its users to be notified of their colleagues' membership in the social network based on their email or phone contact list.	4	1	4	2	11
Introducing peer researchers with the same background	ASNS should introduce researchers with similar research backgrounds to each other based on information entered by users.	1	3	4	2	9
Share information, knowledge, and experiences	The sharing of knowledge and experiences of researchers by the ASNS is critical and considered a requirement for the ASNS's success.	0	2	5	3	10
Exchange opinions individually and in groups	Dissemination of health researchers' in-network opinions and posts to individuals and groups will improve the quality of research.	0	0	5	3	8
Online meetings and session	Although there are several specialized software for online meetings, it seems that the integration and access to more features, such as online meetings, can increase researchers' desire to use ASNSs.	3	1	3	4	11
Disseminating the information by sending posts	Health researchers tended to publish their advertisements, comments, requests, and requirements by sending a post.	3	2	3	4	12
Managing scientific publications						
Scientometrics	Researchers like to use this ASNS to monitor their citations.	2	3	5	4	14
Altmetrics	This network feature shows how many times a document has been downloaded or read, And this can show how important and practical it is.	2	3	5	4	14
Drawing of researchers' network	Showing researcher followers and followers has many advantages for other users. This allows them to get to know another person working in a common field.	2	1	2	1	6
Publish and receive articles	The possibility of publishing and receiving articles is one of the essential parts of ASNS based on health researchers' views.	3	4	5	3	15
Find the best journals for the manuscripts	The ASNS can introduce appropriate journals to users based on user profile data.	3	0	0	1	4
Enhanced the process of conducting research						
Projects management of research	In some cases, health researchers have several responsibilities other than research, including education, patient care, and executive activities. Therefore, providing services via ASNS to manage research projects is very helpful.	0	1	2	3	6
Find help to solve research problems	Various challenges, such as financial, administrative, property rights, laws, and access to protocols, are treated by health researchers during the research stage. As a facilitator, ASNS can play an essential role in solving these problems.	5	2	4	4	15

Theme and subthemes	Description	PhD student (n=6)	Postdoctoral researcher (n=4)	Faculty member (n=6)	Research manager (n=7)	All (n=23)
Providing educational materials	Given that the research is based on innovations and problem solving, new research techniques and methods can be made available to researchers through the ASNS.	3	2	1	1	7
Electronic and reminder calendar (digital Secretary)	Time management is one of the basic principles of research success; ASNS can play an essential role in managing researchers' time by providing tools such as electronic calendars and reminders.	0	1	1	0	2
Contact other researchers	Creating different communication platforms by the network and facilitating communication between researchers promotes cooperation.	5	3	5	4	17
To be informative						
Introducing people with special skills	When a researcher needs an expert with unique skills, the ASNS should support them with search tools.	4	2	3	1	10
Introducing special research services	Researchers have different knowledge and skills. Sometimes they need laboratory services that are not available in their work environment. The research manager can inform you about the services available in a laboratory and share them with other researchers using an ASNS.	1	0	0	4	5
Introducing job or collaboration opportunities	Health researchers can use ASNS to find jobs or collaboration opportunities in research projects.	1	3	4	0	8
Introducing study opportunities	Providing study opportunities on ASNS and creating transparency can lead to a better selection of candidates.	0	2	4	3	9
Introducing ongoing or completed similar projects	Searching for similar research projects and preventing duplicate works can provide good opportunities for collaboration between researchers.	0	3	5	1	9
Introducing the activities of selected researchers	The following possibility of the researcher's activities allows users to obtain information about the latest scientific achievements.	5	2	4	3	14
Introducing scientific and research events	Health researchers can use ASNSs to find out the news related to scientific and research events.	1	3	5	4	13
Sharing and trading laboratory materials and equipment						
Facilitate to exchange or sale of research products	Researchers can share, exchange, or sell their research products on ASNSs.	1	2	2	5	10
Facilitate to sharing of laboratory equipment	By creating the necessary content for sharing laboratory equipment, the ASNS can increase this laboratory equipment's efficiency and provide the opportunity for cooperation between research centers.	4	3	4	5	16
Facilitate to sale and exchange of laboratory materials	In addition to reducing research costs, the exchange of laboratory materials among researchers accelerates research processes.	4	3	3	4	15

Table 2. Technical expectations of health researchers from academic social network sites.

Theme and subthemes	Description	PhD student (n=6)	Postdoctoral researcher (n=4)	Faculty member (n=6)	Research manager (n=7)	All (n=23)
User management capabilities						
Profile management	Creating a user profile is the main feature of academic social networks.	5	3	5	6	19
Share different types of files	Due to the production of knowledge in various formats by health researchers and the communication between them, sharing files in different formats is one of the essential features of the ASNS.	4	4	5	3	16
Automatic in all functions	The automation of ASNS in different functions, such as informing and introducing colleagues, is one of these networks' main strategies to attract researchers.	2	1	1	3	7
Feedbacks management	Posting comments on the ASNS about the researcher's activities in various forms such as text, confirm, and like can improve the researcher's activities. However, the user must manage the type and content of this feedback.	1	0	0	2	3
High security and privacy						
Privacy protection	The most critical concern for health researchers in using ASNS is to protect their privacy.	2	4	5	5	16
High security	Health researchers have become more sensitive to academic and social networks' security.	2	4	4	6	16
Dedicated environment	ASNS' environment should be dedicated to researchers.	0	1	0	3	4
Reliable environment	Researchers need a reliable social media environment to share scientific discussions and share their social media views.	0	2	2	3	7
User friendly						
Understandable	The use of abstract and incomprehensible words in ASNS makes the social network unacceptable for researchers.	2	1	5	3	11
Simplicity	Researchers should simply be able to use ASNS.	2	2	3	1	7
Entertaining	The art of embedding gamification in the academic social network makes researchers more inclined to use this in their daily activities.	3	1	1	1	6
Attractive	Observance of clarity and transparency, visual appeal, color selection, and visuals are criteria for making the ASNS more attractive.	3	2	2	4	11
Customization feature	ASNS does not have wholly fixed characteristics, but a number of its features should be changeable based on the choice and needs of the researcher.	1	2	5	3	11
Other technical features						
Use of cloud store	Today, cloud storage has become an essential requirement for researchers.	2	1	1	1	5
Run on different platforms (responsive)	In the design process of a social network, all pages must be visible on all devices with similar content, design, and performance.	4	3	4	5	16
Lack of error, nonaccessibility, and downtime	The high rate of unavailability of the ASNS can reduce its users.	0	1	1	2	4
Comprehensive	The comprehensiveness and non-allocation of the site to a group of researchers lead to researchers' rapid growth.	1	1	0	0	2

Theme and subthemes	Description	PhD student (n=6)	Postdoctoral researcher (n=4)	Faculty member (n=6)	Research manager (n=7)	All (n=23)
Multilingualism	The multilingual aspect of ASNSs, in addition to understanding, can make a social network more trustworthy and more inclusive.	4	1	1	2	8
Connect with other social networks	Communicating and retrieving information from other academic social networks, in addition to saving time, can make social networks more attractive.	1	2	5	5	13
Regular update	In addition to updating security and information technology, ASNS should create new academic social network functions.	3	1	2	1	7

Technical Features

The technical features are related to the social network's design, language, and the databases and infrastructure used to implement the social network. Based on the views of the study, participants' technical features consist of 4 main themes and 20 subthemes. However, the most important subtheme based on participant study views was profile management, but their concerns about security and privacy were considerable (Table 2).

Discussion

Principal Findings

This study aimed to identify Iranian health researchers' expectations for ASNS from the perspective of a low-income country.

One of the most important expectations of ASNSs was to create a platform for communication and to strengthen researchers' team activities. In line with this study's results, Salahshour [29] showed that 54% of researchers use ASNSs to find colleagues, and 75% of them use ASNS to communicate. Krause [7] argued that in addition to creating intraorganizational communication, ASNSs should facilitate communication and the sharing of resources between scientists. Manca [30] also considers the most important task of ASNSs as establishing a relationship between researchers in the same field. Given that the study participants were researchers from a low-income country, they may have felt a greater need to connect and collaborate with other researchers in high-income countries.

Another functional expectation from ASNSs was the management of publication researchers' work. In line with the results of this study, several studies confirm that publishing management is an essential duty of ASNSs [31-33]. Salahshour [29] also found that 67 percent of users use ASNSs to improve citations and scientific advances. Weber [34] attributes researchers uploading the research results to ASNSs due to increased citations and establishing cooperation and communication between researchers [34]. However, Bonaiuti [9] attributes this behavior to the possibility of receiving feedback and the ease of loading articles in the ASNSs [9]. Because scientometric and altimetric indicators are among the main criteria for evaluating and ranking researchers, health researchers use ASNSs as a tool to display their articles. They try to improve the desired indicators by making their works available. Completing the profile correctly and updating the uploaded items plays a vital role in the researcher being seen by other colleagues. In addition to upgrading the altimetric rankings, ASNSs can improve the number of citations by creating communication capacities and collaborating with traditional metrics.

The researchers' third functional feature involved in the study was to help facilitate research and solve research projects by ASNSs. In a prior study, 56 percent of researchers said their goal for being a member of an ASNS was to improve research quality and learning [29]. In this regard, Espinoza [35] acknowledges that by creating communication, collaboration, and networking platforms, ASNSs support researchers and academics. Various studies have shown that ASNSs, in addition to their tools for communication, collaboration, question and answer, specialized discussion groups, and ability to introduce researchers with the same background, can support researchers and improve research quality [9,34,36]. The advantages of cooperation between researchers include reducing researchers' workload, regulating the activities of researchers based on expertise and skills, increasing the credibility and quality of research, increasing the number of studies, and increasing the productivity and efficiency of researchers. By creating a platform for communication and cooperation on the one hand and maintaining individuals' privacy, on the other hand, ASNSs provide the foundation for the cooperation and facilitation of research. One of the future challenges of ASNSs seems to be managing collaborations between researchers, managing collaboration requests, and protecting researchers' privacy.

Being informative is an essential expectation for ASNSs. Researchers believe, given their interest, an ASNS should automatically inform them of study opportunities, suitable jobs, and cooperation suggestions. Findings from Dermentzi [37] show that one of the purposes of using the search tool in ASNSs is to obtain information. He acknowledges that these sites must collect and process the information required by their users. Another study emphasizes that the researcher should use the ASNS to identify the researchers and create a cooperation network [38]. Meishar [33] stated in addition to finding information, researchers can use these sites to identify new research trends from leading researchers in various fields. The capacity to be informative via different avenues is one of the advantages of ASNS; however, the entry of newly requested and unrelated information by the ASNS into the email and the researcher's account can be considered a weakness for the ASNS and cause the user to leave the ASNS. Customization,

artificial intelligence algorithms, and user engagement in information acquisition can prevent this challenge and improve the quality of ASNS-related information.

The fifth practical feature considered by Iranian health researchers participating in the study was the possibility of facilitating the sharing and trading of laboratory materials and equipment amongst researchers. Bonaiuti [9] acknowledges that researchers can meet their needs using public posts on social media or specialized groups, which helps users of that social network find or share research resources. In the Salahshour study [29], 73% of researchers used ASNS to find material related to their research [29]. The existence of specialized groups in the ASNSs can be an effective solution for sharing laboratory materials and equipment and bolstering effective communication [9]. Given the situation in Iran and the sanctions imposed on the one hand [39] and poor economic conditions, on the other hand, this user expectation seems reasonable. Users can share the features of their laboratory materials and equipment and share their resources with other researchers. In addition to economic savings, this practice can increase research centers' efficiency and strengthen cooperation between researchers.

Conclusions

This study aimed to identify the expectations of health researchers from ASNSs. These expectations were divided into functional and technical characteristics. Functional characteristics were related to different research processes, and researchers used these features to increase the speed and quality of their research. In this category, they expected ASNSs to facilitate communication and inform them about various research fields. Moreover, some researchers expected ASNSs to enhance the process of conducting research and help in sharing and trading laboratory materials and equipment. Managing scientific publications is a functional characteristic that includes improving and managing scientometrics and altmetrics, introducing related journals, publishing the researcher's work, raising awareness regarding the scientific ranking of other researchers, and presenting a cooperative network.

Participants' expectations of ASNSs regarding technical characteristics included user management capabilities, high security and privacy, user-friendly, and other technical features. In addition to not meeting the user's expectations of ASNS, it is abandoned by researchers in some cases due to the lack of attention by programmers to users' opinions in the design of ASNS.

Acknowledgments

Thanks to the Vice-Chancellor for Research and Technology of Tabriz University of Medical Sciences. All participants in this study are also sincerely appreciated.

Authors' Contributions

RF, MD, and SA contributed to the study design. MD and RF completed the data collection. MK, SA, and MD conducted the data analysis. RF, SA, and BM supervised the study. MK, MD, and BM were involved in the manuscript writing. All authors contributed critical revisions for important intellectual content.

Conflicts of Interest

None declared.

References

1. Gruzd A. Non-academic and academic social networking sites for online scholarly communities. Social Media for Academics 2012:21-37. [doi: 10.1016/b978-1-84334-681-4.50002-5]
2. Bik HM, Goldstein MC. An Introduction to Social Media for Scientists. PLoS Biol 2013 Apr 23;11(4):e1001535. [doi: 10.1371/journal.pbio.1001535]
3. Khvatova T, Dushina S, Nikolaenko G. Do the Online Activities of Scientists in Social Professional Networks Influence their Academic Achievements? European Conference on Management, Leadership & Governance; : Academic Conferences International Limited. In: ECMLG 2017 13th European Conference on Management, Leadership and Governance. 2017 Presented at: ECMLG 2017 13th European Conference on Management, Leadership and Governance; 2017; Valletta p. 217-226 URL: https://www.academia.edu/35656146/Do_the_Online_Activities_of_Scientists_in_Social_Professional_Networks_Influence_their_Academic_Achievements
4. Kulathuramaiyer N. How Social Networks will Change Research. The IPSI BgD Transactions on Internet Research 2016:15-20 [FREE Full text]
5. Fiesler C, Proferes N. "Participant" Perceptions of Twitter Research Ethics. Social Media plus Society 2018 Mar 10;4(1):205630511876336. [doi: 10.1177/2056305118763366]
6. Dehghani M, Akhondzadeh S, Mesgarpour B, Ferdousi R. A Tool to Reduce the Problems of Iranian Health Researchers. ijph 2020 Dec 09;49(12):2441-2442. [doi: 10.18502/ijph.v49i12.4840]
7. de Rosa A, Bocci E, Dryjanska L, Borrelli F. The role of academic social networking in the dissemination of the social representations literature. 10h International Technology, Education and Development Conference, INTED Proceedings

(Valencia, SPAIN, th of March,). 2016 Presented at: 10h International Technology, Education and Development Conference; 2016; Valencia p. 7-9. [doi: 10.21125/inted.2016.1237]
8. Krause J. Tracking references with social media tools: organizing what you've read or want to read. In: Social Media for Academicslsevier. Amsterdam: Chandos Publishing; 2012:85-104.
9. Bonaiuti G. Academic Social Networks: How the web is changing our way to make and communicate researches. Research on education and media 2015;7(2):3-14. [doi: 10.1515/rem-2015-0010]
10. Murray M. Analysis of a scholarly social networking site: The case of the dormant user. SAIS 2014 Proceedings 2014;21(3):1-6 [FREE Full text]
11. Manca S. ResearchGate and Academia.edu as networked socio-technical systems for scholarly communication: a literature review. Research in Learning Technology 2018 Feb 20;26(1):1-16. [doi: 10.25304/rlt.v26.2008]
12. Liu D, Kirschner PA, Karpinski AC. A meta-analysis of the relationship of academic performance and Social Network Site use among adolescents and young adults. Computers in Human Behavior 2017 Dec;77:148-157. [doi: 10.1016/j.chb.2017.08.039]
13. Salvation M, Adzharuddin N. The influence of social network sites (SNS) upon academic performance of Malaysian students. International Journal of Humanities and Social Science 2014;4(10):131-138 [FREE Full text]
14. Ortega J. Disciplinary differences in the use of academic social networking sites. Online Information Review 2015;39(4):520-536. [doi: 10.1108/oir-03-2015-0093]
15. Ghorbani N, Momeni M, Ghorbani R, Babalhavaeji F. A Study on the Presence of Iranian Researchers in Academic Social Networks: A Case Study on the Faculty Members of Semnan University Of Medical Sciences, Iran. Health Information Management 2018:260-266.
16. Mozaffari A, Rastegari B. Evaluation of The Iranian Users Trust to Privacy in Linkedin. Media Studies 2015;10(28):79-94 [FREE Full text]
17. Ghazimirsaeed S, Papi A, Ramezani A, YektaKooshali M, RamezaniPakpourLangroudi F. Evaluation Altmetric Indicators of Iranian Medical Universities in Academic Social Networks: ResearchGate and Academia.edu. Quarterly Knowledge and Information Management Journal 2018;4(1):23-30 [FREE Full text]
18. Jeng W, He D, Jiang J. User participation in an academic social networking service: A survey of open group users on Mendeley. J Assn Inf Sci Tec 2014 Jun 17;66(5):890-904. [doi: 10.1002/asi.23225]
19. Dehghani M, Akhondzadeh A, Mesgarpour B, Ferdousi R. Design and Implementation of a Social Network for Laboratory Researchers. ircmj 2020 Oct 31;22(10):1-6. [doi: 10.32592/ircmj.2020.22.10.103]
20. Azami-Aghdash S, Ghojazadeh M, Aghaei M, Naghavi-Behzad M, Asgarlo Z. Perspective of patients, patients' families, and healthcare providers towards designing and delivering hospice care services in a middle income Country. Indian J Palliat Care 2015;21(3):341-348. [doi: 10.4103/0973-1075.164898]
21. Ghojazadeh M, Azami-Aghdash S, Sohrab-Navi Z, Kolahdouzan K. Cardiovascular patients' experiences of living with pacemaker: Qualitative study. ARYA Atheroscler 2015 Sep;11(5):281-288 [FREE Full text] [Medline: 26715933]
22. Englander M. The phenomenological method in qualitative psychology and psychiatry. International Journal of Qualitative Studies on Health and Well-being 2016 Mar 09;11(1):1-11. [doi: 10.3402/qhw.v11.30682]
23. Byrne M. Sampling for qualitative research. AORN Journal 2001 Feb 01;73(2):494-498. [doi: 10.1016/s0001-2092(06)61990-x]
24. Cleary M, Horsfall J, Hayter M. Data collection and sampling in qualitative research: does size matter? J Adv Nurs 2014 Jan 23;70(3):473-475. [doi: 10.1111/jan.12163]
25. Higginbottom GMA. Sampling issues in qualitative research. Nurse Researcher 2004 Jul;12(1):7-19. [doi: 10.7748/nr2004.07.12.1.7.c5927]
26. Grbich C. Qualitative data analysis: An introduction. Thousand Oaks, CA: Sage; 2007:27-39.
27. Hsieh H, Shannon SE. Three Approaches to Qualitative Content Analysis. Qual Health Res 2016 Jul 01;15(9):1277-1288. [doi: 10.1177/1049732305276687]
28. Ope C, Ziebland S, Mays N. Analysing qualitative data. British Medical Journal 2000;320(7227):114-116. [doi: 10.1002/9780470750841.ch7]
29. Salahshour M, Dahlan H, Iahad N. A case of academic social networking sites usage in Malaysia: drivers, benefits, and barriers. International Journal of Information Technologies and Systems Approach (IJITSA) 2016;9(2):88-99. [doi: 10.4018/ijitsa.2016070106]
30. Manca S, Raffaghelli J. Towards a multilevel framework for analysing academic social network sites: A networked socio-Technical perspective. In: 4th European Conference on Social Media–ECSM. 2017 Presented at: 4th European Conference on Social Media–ECSM; 2017; Poland p. 193-201 URL: https://www.academia.edu/33555161/Towards_a_Multilevel_Framework_for_Analysing_Academic_Social_Network_Sites_A_Networked_Socio_Technical_Perspective
31. Aleryani A, Mofleh H, Alariki S. The Usage of Academic Social Network Sites by Researchers in Developing Countries: Opportunities and Challenges. Journal of Information Technology and Networking 2017:49-60 [FREE Full text]
32. Bardakcı S, Arslan Ö, Ünver TK. How scholars use academic social networking services. Information Development 2017 May 24;34(4):334-345. [doi: 10.1177/0266666917712108]

33. Meishar-Tal H, Pieterse E. Why Do Academics Use Academic Social Networking Sites? IRRODL 2017 Feb 28;18(1):1-23. [doi: 10.19173/irrodl.v18i1.2643]
34. Weber GM, Barnett W, Conlon M, Eichmann D, Kibbe W, Falk-Krzesinski H, et al. Direct2Experts: a pilot national network to demonstrate interoperability among research-networking platforms. Journal of the American Medical Informatics Association 2011 Dec 01;18(2):157-160. [doi: 10.1136/amiajnl-2011-000200]
35. Espinoza VF, Caicedo BC. Academic social networking sites: A comparative analysis of their services and tools. IConference 2015 Proceedings 2015:1-6. [doi: 10.1007/978-1-137-33497-8_3]
36. El-Berry DK. Awareness and Use of Academic Social Networking Sites by the Academic Staff at the South Valley University in Egypt. JLIS 2015;3(2):115-132. [doi: 10.15640/jlis.v3n2a7]
37. Dermentzi E, Papagiannidis S. Academics' intention to adopt online technologies for public engagement. INTR 2018 Feb 06;28(1):191-212. [doi: 10.1108/intr-10-2016-0302]
38. Zerhouni EA. Translational and Clinical Science — Time for a New Vision. N Engl J Med 2005 Oct 13;353(15):1621-1623. [doi: 10.1056/nejmsb053723]
39. Dehghani M, Mesgarpour B, Akhondzadeh S, Azami-Aghdash S, Ferdousi R. How the US Sanctions Are Affecting the Health Research System in Iran? Arch Iran Med 2021 Feb 01;24(2):101-106. [doi: 10.34172/aim.2021.15] [Medline: 33636976]

Abbreviations

ASNS: academic social network site

Edited by R Kukafka; submitted 28.09.20; peer-reviewed by G Weber; comments to author 23.10.20; revised version received 29.11.20; accepted 24.05.21; published 07.12.21.

Please cite as:
Dehghani M, Kahouei M, Akhondzadeh S, Mesgarpour B, Ferdousi R
Expectations of Health Researchers From Academic Social Network Sites: Qualitative Study
J Med Internet Res 2021;23(12):e24643
URL: https://www.jmir.org/2021/12/e24643
doi:10.2196/24643
PMID:34878993

Original Paper

Effects of Social Networking Service (SNS) Addiction on Mental Health Status in Chinese University Students: Structural Equation Modeling Approach Using a Cross-sectional Online Survey

Tingxuan Wang[1], MPhil; Janet Y H Wong[1], PhD; Man Ping Wang[1], PhD; Amanda Chiu Yin Li[2]; Sang Suk Kim[3], PhD; Jung Jae Lee[1], PhD

[1]School of Nursing, The University of Hong Kong, Hong Kong, China

[2]Department of Medicine, The University of Hong Kong, Hong Kong, China

[3]Red Cross College of Nursing, Chung-Ang University, Seoul, Republic of Korea

Corresponding Author:
Jung Jae Lee, PhD
School of Nursing
The University of Hong Kong
4/F William MW Mong Block Building
21 Sassoon Rd, Pokfulam
Hong Kong
China
Phone: 852 39176971
Fax: 852 28726079
Email: leejay@hku.hk

Abstract

Background: Although social networking services (SNSs) have become popular among young people, problematic SNS use has also increased. However, little is known about SNS addiction and its association with SNS use patterns and mental health status.

Objective: This study aims to test the mediating role of SNS addiction between SNS use patterns and mental health status among Chinese university students in Hong Kong (HK).

Methods: An online cross-sectional survey was conducted using a convenience sampling method. In total, 533 university students (323 [66.9%] female, mean age [SD]=20.87 [2.68] years) were recruited from February to March 2019. Multiple linear regression was used to assess the association between SNS use and SNS addiction. Structural equation modeling (SEM) was performed to examine the pathways and associations among SNS use, SNS addiction, psychosocial status, and mental health status (including anxiety and depressive symptoms).

Results: A longer time spent on SNSs per day (>3 h), a longer time spent on each SNS access (≥31 min), a higher frequency of SNS access (≤every 30 min), a longer duration of SNS use before sleeping (≥61 min), and a shorter duration from waking to first SNS use (≤5 min) were significantly associated with a higher level of SNS addiction (adjusted beta [aβ]=6.03, 95% CI 4.66-7.40; aβ=4.99, 95% CI 3.14-6.83; aβ=5.89, 95% CI 4.14-7.64; aβ=5.92, 95% CI 4.19-7.65; and aβ=3.27, 95% CI 1.73-4.82, respectively). SEM showed a significant mediating effect of SNS addiction in the relationship between SNS use and psychosocial status, and mental health status, including an indirect effect (β=0.63, 95% CI 0.37-0.93) and the total effect (β=0.44, 95% CI 0.19-0.72), while the direct effect was insignificant (β=–0.19, 95% CI –0.49 to 0.08).

Conclusions: SNS use patterns were associated with SNS addiction, and SNS addiction mediated the associations between SNS use, psychosocial status, and mental health status of Chinese university students in HK. The findings suggest that screening for and addressing excessive SNS use are needed to prevent SNS addiction and mental distress among young people.

(J Med Internet Res 2021;23(12):e26733) doi:10.2196/26733

KEYWORDS

social networking service; SNS; addiction; depression; anxiety; psychosocial status; youth; mental health

Introduction

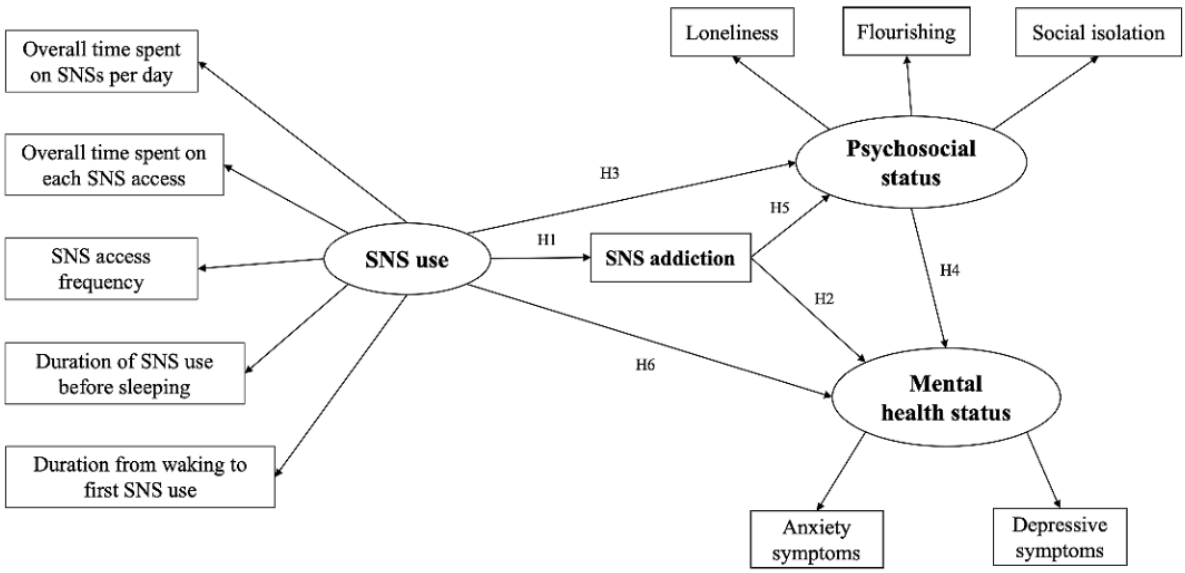
More than 1 billion people worldwide regularly use social networking services (SNSs), such as Facebook, Twitter, and Instagram, which are virtual communities where users interact and build online and real-life relationships [1-3]. Similar to the global trend, the number of SNS users has substantially increased from 68% in 2014 to 83% in 2018 in Hong Kong (HK) [4] which has 1 of the highest penetration rates of the internet and e-devices [5,6]. Compared with other age groups, younger age groups more popularly use SNSs, and 95% of young people in HK, aged from 10 to 24 years, used SNSs for an average of 17.7 h/week (the longest SNS use than other age groups) in 2018 [4]. As SNS use has gained popularity, problematic SNS use, such as excessive SNS use, which is a relatively long duration or frequent use of SNSs [7], has also increased among young people [8,9]. Evidence has shown that excessive SNS use has a negative impact on young people's physical health (eg, headache, postural pain, and eye strain) [7,10], real-life social relationships [11], and academic performance [12,13]. Moreover, although SNS platforms provide more pervasive environments for individuals to share their feelings and follow other interesting friends, previous studies have reported that excessive SNS use shows a strong association with psychological distress, including depression and anxiety [13-16], such as feeling frustrated or isolated from their real life [17] and feeling vulnerable to the perceptions of others [18].

Excessive SNS use has shown significant association with addictive behaviors [19,20], including feeling constantly anxious or uneasy about one's SNS accounts and having strong urges to use SNSs. SNS addiction is thus defined as the excessive use of SNSs that affects one's daily activities, interpersonal relationships, and mental health [21]. Although SNS addiction has not been specifically identified by the *Diagnostic and Statistical Manual 5th Edition* (DSM-5) or the *International Classification of Diseases 11th Revision* (ICD-11), SNS addiction can be considered as 1 form of behavior addiction, reflecting excessive and compulsive symptoms [22,23], similar to internet gaming disorder. However, SNS addiction can be differentiated from other behavior addictions as users focus on interpersonal networking through SNSs. According to gratification theory, users who gain enjoyment and happiness through SNS platforms would continue to spend even more time using SNSs in pursuit of more positive gratification [24]. Therefore, SNS addiction may result from gratification attained through online surfing and through virtual and real-life interpersonal relationships. Meanwhile, recent studies have found that 29% of university students have SNS addiction and that SNS addiction is associated with anxiety, depression, and mania symptoms, similar to prior studies that investigated excessive SNS use [25,26].

Although recent studies have identified associations between excessive SNS use, SNS addiction, and mental health status, it remains unknown whether there is a clear pathway from excessive SNS use to SNS addiction, thus influencing users' mental health (ie, pathway). As SNS use patterns can be differentiated by their social contexts (eg, different cultural values in interpersonal relationships on SNSs [27]), excessive SNS use and SNS addiction can be influenced by the users' psychosocial status (ie, the influence of social factors on one's mind [28]), such as the influence of social isolation and loneliness precipitating SNS use [29-31], while conversely influencing their mental health. Moreover, the possible mediating effect of SNS addiction on the pathway has been deduced as SNS addiction is also associated with SNS use patterns, psychosocial status, and mental health status [32-34]. Therefore, this study aimed to develop a structural equation modeling (SEM) approach to testing a hypothetical model and the potential mediating role of SNS addiction among Chinese university students in HK.

Based on the existing literature and knowledge gap, we developed the following hypotheses for this study and a hypothetical model (Figure 1):

Figure 1. Conceptual model and hypotheses. Ovals represent unobserved latent variables. Rectangles represent observed measured variables. SNS: social networking service.



- Hypothesis 1 (H1): SNS use is positively associated with SNS addiction.
- Hypothesis 2 (H2): SNS addiction is negatively associated with mental health status.
- Hypothesis 3 (H3): SNS use is negatively associated with psychosocial status.
- Hypothesis 4 (H4): Psychological factors are positively associated with mental health status.
- Hypothesis 5 (H5): SNS addiction is negatively associated with psychosocial status.
- Hypothesis 6 (H6): SNS use is negatively associated with mental health status.

Methods

Study Design and Sampling

A cross-sectional study design was adopted. An online survey link (Google Form) and a QR code printed on survey invitation flyers were distributed in busy public areas (eg, the entrance of library and cafeteria) within 2 major public university campuses in HK from February to March 2019. The online survey website included the study aim, respondents' rights, and participation incentive (50 respondents were randomly selected through a lucky draw for a HK $25 [around US $3.2] cash voucher). The respondents were also invited to freely share the survey link with their peers from any HK university (ie, convenience and snowball sampling). The inclusion criteria were enrolled university students residing in HK with at least 1 SNS account. Exchange students from outside HK and respondents who did not answer more than 10 survey questions were excluded. To prevent multiple responses and confirm their identity as university students, each respondent's mobile phone number and university email address were collected.

Measurements

Demographics and SNS Use

Demographic characteristics, including sex, age, university, degree pursued, and academic performance, were investigated. The respondents' patterns of SNS use (eg, overall time spent on SNS per day, time spent on each SNS access, SNS access frequency, and SNS use duration before sleeping and after waking) were surveyed.

SNS Addiction

SNS addiction was measured using a revised version of the Bergen Facebook Addiction Scale (BFAS), which consists of 6 questions [35,36]. The BFAS presented 0.83 internal consistency and reported convergent validity [35]. The overall score of the BFAS ranged from 6 to 30, with higher scores indicating a higher level of SNS addiction. The term "Facebook" in each item was changed to "SNSs" to include a wider range of SNSs, similar to Andreassen et al [36]. The internal consistency (Cronbach α) of the BFAS in this study was also 0.83.

Mental Health Status

The Patient Health Questionnaire-4 (PHQ-4) was used to screen for psychological symptoms among the respondents. The PHQ-4 consists of 2 domains (ie, depression and anxiety), and each domain has 2 questions [37]. The total score in each domain ranges from 0 to 6. Higher scores in the 2 domains refer to higher levels of depressive and anxiety symptoms, respectively. The internal consistency of the 2 domains was 0.82 and 0.81, respectively, in the original study [37] and 0.79 and 0.86, respectively, in this study.

Psychosocial Status

The respondents' psychosocial status was measured according to their psychological well-being, social isolation, and loneliness. First, the Flourishing Scale was adopted to measure the respondents' subjective psychological well-being [38]. The scale consists of 8 questions with an overall score range from 8 to 56, and higher scores represent higher subjective psychological well-being. The Flourishing Scale showed an internal consistency as 0.87 and convergent validity [38]. The internal consistency in this study was 0.92. Second, we adopted the Lubben Social Network Scale (LSNS-6) consisting of 6 questions to investigate the respondents' social isolation level [39]. The overall score ranges from 0 to 30, and higher scores indicate lower levels of social isolation [39]. The internal consistency of the LSNS-6 was 0.83 in the original study and 0.75 in this study. Lastly, a brief loneliness scale [40] was also adopted to measure the respondents' feelings of loneliness. The scale consists of 3 questions, and the total score ranges from 3 to 9. The internal consistency of the loneliness scale in the original research [40] and this study was 0.72 and 0.89, respectively.

Statistical Analysis

Descriptive analysis was conducted to detail the respondents' information about demographic characteristics, SNS use and addiction, and psychological profile. For continuous variables, the mean and SD were used, and for categorical variables, the frequency and percentage were used for analysis. Multiple linear regression was performed to assess the association between SNS use and SNS addiction. Structural equation modeling (SEM) was performed to examine the pathways from SNS use to mental health status. Multiple imputation was performed for missing data. Total, direct, and indirect effects in the hypothesized model were estimated using the maximum likelihood and bias-corrected (BC) 95% CI by the bootstrapping method with 2000 replications. Fitness indices of the SEM were considered, including the root-mean-square error of approximation (RMSEA; suggested close to or smaller than 0.06), the comparative fit index (CFI; suggested close to or larger than 0.95), the Tucker-Lewis index (TLI; suggested close to or larger than 0.95), and the standardized root-mean-square residual (SRMR; suggested close to or smaller than 0.08) [41,42]. IBM SPSS 24.0 and Amos 26.0 were used for all analyses at a .05 significance level.

Results

Respondents' Characteristics

Of the 533 respondents, 483 (90.6%) who met the eligible criteria were included in this study. The mean age was 20.87 years (SD 2.68), and 66.9% (n=323) of the respondents were female. The majority (459/483, 95.0%) were from University

Grants Committee (UGC)-funded universities (ie, large public universities) and undertook higher diploma or undergraduate courses (466/483, 96.5%). With regard to SNS use, 33.7% (163/483), 42.2% (225/483), and 31.0% (165/483) of the respondents used SNSs for over 1-2 h/day, 6-15 min each time, and every 31-60 min, respectively. In addition, 46.9% (250/483) and 36.0% (192/483) of them used SNSs for 31-60 min before sleeping and within 5 min or less after waking up. The mean scores for anxiety symptoms, depressive symptoms, flourishing, social isolation, and loneliness were 2.86 (SD 1.62), 2.70 (SD 1.58), 38.87 (SD 7.42), 16.90 (SD 5.06), and 5.55 (SD 1.79), respectively. The mean score for SNS addiction was 16.4 (SD 5.03); see Table 1.

Table 1. Demographic characteristics and frequency of SNS[a] use and SNS addiction (N=483).

Variables	n (%)	SNS addiction, mean (SD)	SNS addiction, *P* value[b]
Sex (*P*=.01)			
Male	160 (33.1)	15.49 (5.15)	__[c]
Female	323 (66.9)	16.84 (4.92)	—
University (*P*=.76)			
UGC[d] funded	459 (95.0)	16.41 (0.06)	—
Non-UGC funded	24 (5.0)	16.08 (4.63)	—
Degree of study (*P*=.002)			
Higher diploma and undergraduate	466 (96.5)	16.53 (5.03)	—
Postgraduate	17 (3.5)	12.65 (5.35)	—
Academic performance (*P*<.001)			
First-class honors/quartile 4 or equivalent	84 (17.1)	14.92 (5.03)	—
Second-class honors/quartile 3 or equivalent	192 (39.8)	15.89 (4.90)	—
Second-class honors/quartile 2 or equivalent	118 (24.4)	18.03 (4.53)	—
Third-class honors/quartile 1 or equivalent	20 (4.1)	17.55 (0.90)	—
Missing	69 (14.3)	—	—
Overall time spent on SNS per day (*P*<.001)			
≤1 h	72 (13.5)	12.43 (4.84)	—
>1-2 h	163 (33.7)	16.42 (4.60)	—
>2-3 h	125 (25.9)	18.40 (4.67)	—
>3 h	123 (25.5)	20.29 (4.85)	—
Time spent on each SNS access (*P*<.001)			
5 min or less	122 (22.9)	14.89 (5.05)	—
6-15 min	225 (42.2)	16.10 (4.84)	—
16-30 min	100 (18.8)	17.56 (4.60)	—
31 min or more	36 (6.7)	20.08 (4.94)	—
SNS access frequency (*P*<.001)			
Every 5 min or sooner	30 (5.6)	17.73 (6.73)	—
Every 6-30 min	141 (26.5)	17.85 (4.32)	—
Every 31-60 min	165 (31.0)	16.42 (4.77)	—
Every 61 mins or later	147 (27.7)	14.69 (5.08)	—
Duration of SNS use before sleeping (*P*<.001)			
61 min or more	43 (8.1)	14.03 (5.02)	—
31-60 min	94 (17.6)	16.28 (4.55)	—
6-30 min	250 (46.9)	18.49 (4.28)	—
5 mins or less	74 (13.9)	19.26 (5.32)	—
Duration from waking up to first SNS use (*P*=.02)			
5 min or less	192 (36.0)	16.83 (5.12)	—
6-30 min	178 (33.4)	16.92 (4.72)	—
31-60 min	64 (12.0)	15.77 (4.57)	—
61 min or more	34 (6.4)	14.29 (5.45)	—

Variables	n (%)	SNS addiction, mean (SD)	SNS addiction, *P* value[b]
Psychosocial status and mental health status			
Loneliness	—	5.55 (1.79)	<.001
Flourishing	—	38.87 (7.42)	<.001
Social isolation	—	16.90 (5.06)	<.001
Anxiety symptoms	—	2.86 (1.62)	<.001
Depressive symptoms	—	2.70 (1.58)	<.001

[a]SNS: social networking service.

[b]*P* value for the *t* test (1 or 2 groups) or ANOVA (3 groups or more).

[c]Not applicable.

[d]UGC: University Grants Committee. UGC-funded universities in HK including the University of Hong Kong (350/459, 76.3%), the Chinese University of Hong Kong (14/459, 3.1%), the Hong Kong University of Science and Technology (3/459, 0.6%), the Hong Kong Polytechnic University (81/459, 17.6%), the Education University of Hong Kong (0), the City University of Hong Kong (7/459, 1.5%), the Hong Kong Baptist University (3/459, 0.6%), and Lingnan University (1/459, 0.2%).

Associations of SNS Addiction With SNS Use

SNS addiction was significantly associated with SNS use patterns (Table 2). A longer amount of time on SNSs per day (ie, >3 h), a longer amount of time on SNSs each access (ie, ≥31 min), and a higher frequency of SNS access (ie, less than every 30 min) were associated with a higher level of SNS addiction (adjusted beta [aβ]=6.03, 95% CI 4.66-7.40; aβ=4.99, 95% CI 3.14-6.83; and aβ=5.89, 95% CI 4.14-7.64, respectively). A longer duration of SNS use before sleeping (ie, ≥61 min) and a shorter duration from waking to first SNS use (ie, ≤5 min) were significantly associated with a higher level of SNS addiction (aβ=5.92, 95% CI 4.19-7.65 and aβ=3.27, 95% CI 1.73-4.82, respectively).

Table 2. Linear regression for the association between SNS[a] use and SNS addiction.

Model	Crude		Adjusted[b]	
	β (95% CI)	β	β (95% CI)	β
Overall time spent on SNSs per day				
≤1 h	REF[c]	—[d]	REF	—
>1-2 h	3.35 (2.07-4.64)[e]	0.32	3.24 (1.94-4.54)[e]	0.30
>2-3 h	4.82 (3.47-6.17)[e]	0.42	4.63 (3.26-6.00)[e]	0.40
>3 h	6.23 (4.88-7.58)[e]	0.54	6.03 (4.66-7.40)[e]	0.52
Time spent on each SNS access				
≤5 mins	REF	—	REF	—
6-15 mins	1.21 (0.14-2.28)[f]	0.12	1.08 (–0.00 to 2.16)	0.11
16-30 mins	2.67 (1.38-3.95)[e]	0.21	2.42 (1.07-3.76)[e]	0.19
≥31 mins	5.19 (3.38-7.00)[e]	0.27	4.99 (3.14-6.83)[e]	0.26
SNS access frequency				
≥Every 6 h	REF	REF	REF	REF
Every 3 h	3.34 (1.51-5.17)[e]	0.28	3.29 (1.48-5.11)[e]	0.28
Every 1 h	4.28 (2.52-6.04)[e]	0.40	4.21 (2.47-5.96)[e]	0.40
≤Every 30 min	5.72 (3.96-7.48)[e]	0.54	5.89 (4.14-7.64)[e]	0.56
Duration of SNS use before sleeping				
≤5 min	REF	—	REF	—
6-30 min	2.92 (1.81-4.02)[e]	0.29	2.83 (1.69-3.98)[e]	0.28
31-60 min	5.12 (3.79-6.46)[e]	0.40	5.10 (3.70-6.51)[e]	0.40
≥61 min	5.89 (4.20-7.58)[e]	0.33	5.92 (4.19-7.65)[e]	0.34
Duration from waking to first SNS use				
≥61 min	REF		REF	
31-60 min	2.17 (0.33-4.02)[f]	0.15	2.15 (0.32-3.99)[f]	0.15
6-30 min	3.33 (1.76-4.90)[e]	0.32	3.28 (1.73-4.84)[e]	0.31
≤5 min	3.24 (1.69-4.80)[e]	0.32	3.27 (1.73-4.82)[e]	0.32

[a]SNS: social networking service.

[b]Adjusted for sex, age, academic performance.

[c]REF: reference group.

[d]Not applicable.

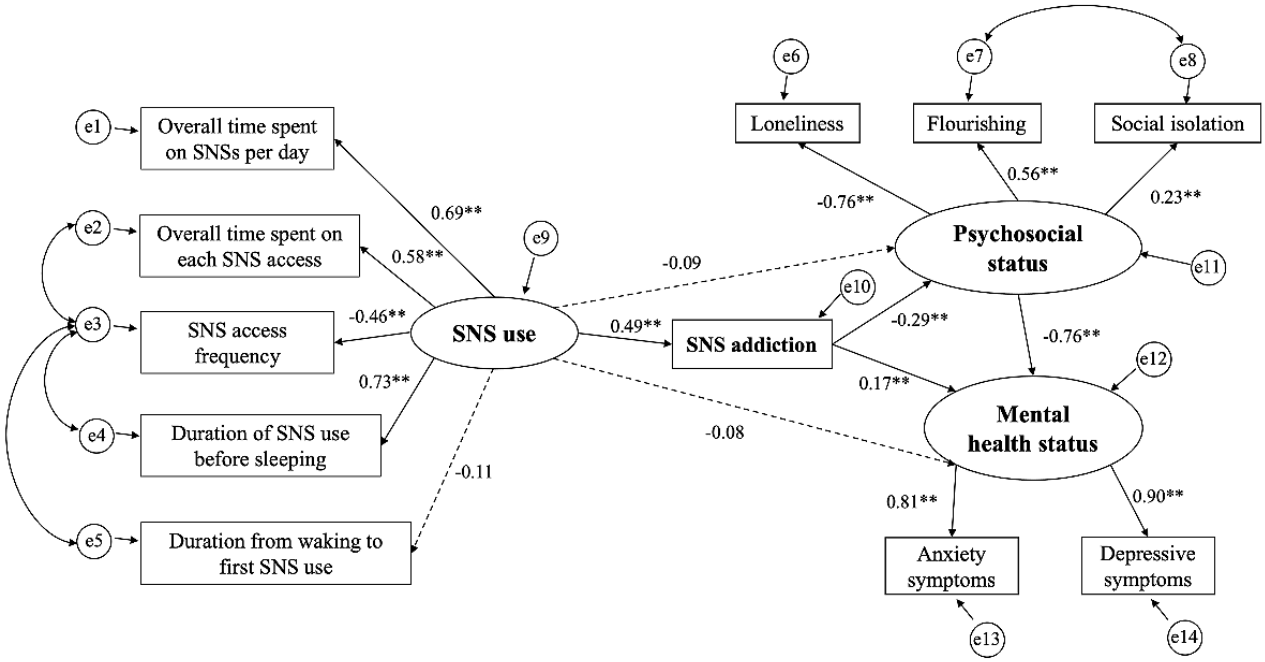
[e]*P*<.001.

[f]*P*<.05

Mediating Effects of SNS Addiction Between SNS Use and Mental Health Status

The hypothesized model showed good fitness indices, with X^2/df=2.16, RMSEA=0.05, CFI=0.97, TLI=0.95, and SRMR=0.04. Results from bootstrapping showed that SNS addiction has a significant mediating effect on the relationship between SNS use and psychosocial status, and mental health status (Figure 2).

Figure 2. Standardized regression coefficient (β) of all variables in the hypothesized model. Ovals represent unobserved latent variables. Rectangles represent observed measured variables. Values are standardized path coefficients. ***P*<.01. SNS: social networking service.



H1 (SNS use and SNS addiction; β=0.49, 95% CI 0.39-0.58), H2 (SNS addiction and mental health status; β=0.17, 95% CI 0.06-0.27), H4 (psychosocial status and mental health status; β=–0.76, 95% CI –0.88 to –0.64), and H5 (SNS addiction and psychosocial status; β=–0.29, 95% CI –0.42 to –0.16) were supported. In contrast, H3 (SNS use and psychosocial status; β=–0.09, 95% CI –0.24 to 0.06) and H6 (SNS use and mental health status; β=–0.08, 95% CI –0.21 to 0.03) were not supported (Table 3).

Table 3. Standardized coefficients of raised hypotheses in the hypothesized model.

Hypothesis	β	BC[a] 95% CI	*P* value
H1: SNS[b] use → SNS addiction	0.49	0.39-0.58	.001
H2: SNS addiction → mental health status	0.17	0.06-0.27	.004
H3: SNS use → psychosocial status	–0.09	–0.24 to 0.06	.21
H4: Psychosocial status → mental health status	–0.76	–0.88 to –0.64	.002
H5: SNS addiction → psychosocial status	–0.29	–0.42 to –0.16	.001
H6: SNS use → mental health status	–0.08	–0.21 to 0.03	.16

[a]BC: bias-corrected.

[b]SNS: social networking service.

Three indirect pathways between SNS use and mental health status in SEM were investigated (Tables 4 and 5). The indirect effect of SNS addiction was significant on the pathway of SNS use to mental health status (β=0.20, 95% CI 0.08-0.34), but the indirect effect of psychosocial status was insignificant on the pathway (β=0.17, 95% CI –0.10 to 0.47). The indirect effect mediated by SNS addiction and psychosocial status in the pathway also showed a significant impact (β=0.26, 95% CI 0.14-0.44), and this pathway showed a stronger correlation than the pathway with SNS addiction alone as the mediator. The total indirect effect between SNS use and mental health status was significant (β=0.63, 95% CI 0.37-0.93). The direct effect of SNS addiction and psychosocial status on mental health status was insignificant (β=–0.19, 95% CI –0.49 to 0.08). Furthermore, the total effect of SNS addiction and psychosocial status that mediated the path from SNS use to mental health status was significant (β=0.44, 95% CI 0.19-0.72).

Table 4. Bootstrapping analyses to examine the indirect effect estimates of pathways in the hypothesized model.

Indirect effect	β	Product of coefficients		BC[a] 95% CI	*P* value
		SE	Z		
SNS[b] use → SNS addiction → mental health status	0.20	0.07	3.02	0.08-0.34	.003
SNS use → psychosocial status → mental health status	0.17	0.14	1.20	–0.10 to 0.47	.21
SNS use → SNS addiction → psychosocial status → mental health status	0.26	0.08	3.40	0.14-0.44	.001

[a]BC: bias-corrected.

[b]SNS: social networking service.

Table 5. Total, direct, and indirect effects of mediation analysis in the model.

Effect	β	Product of coefficients		BC[a] 95% CI	*P* value
		SE	Z		
Total	0.44	0.14	3.22	0.19-0.72	.002
Direct	–0.19	0.15	1.34	–0.49 to 0.08	.17
Indirect	0.63	0.14	4.52	0.37-0.93	.001

[a]BC: bias-corrected.

Discussion

Principal Findings

This study found that longer and more frequent use of SNSs is significantly associated with SNS addiction. This study also identified that SNS addiction and psychosocial status significantly and positively mediate the relationship between SNS use and mental health status (anxiety and depressive symptoms) in Chinese university students in HK.

SNS Addiction–Related Factors

SNS addiction significantly differed across the respondents' demographics in this study. We found that females have a higher level of SNS addiction than males, corresponding with another study from HK [43]. In contrast, several studies from Europe have reported that males have a higher SNS (Twitter) addiction level and SNS obsession than females [44-46]. Although there are limited studies investigating the sex differences in SNS addiction, the distinct findings in the HK studies may be due to sociocultural influences on SNS use and addiction. Chinese females who are influenced by a collectivism-dominant culture tend to be introverted and gravitate toward indirect communication styles [47]. Accordingly, they prefer to use SNSs as a way to make emotional connections and maintain their existing interpersonal relationships [48]. They also tend to spend more time on SNSs than males [47,48]; thus they are more exposed to the risk of SNS addiction. An in-depth qualitative study and meta-analysis would be helpful to uncover the extent of sociocultural influences on SNS addiction.

We also found that SNS addiction is significantly higher among undergraduate students and those with poor academic performance, consistent with previous findings on academic performance [49,50] and undergraduate studies [51]. As time spent on SNSs for entertainment is a negative predictor of time spent studying [52] and academic performance, such as grade point averages [51], excessive and addictive SNS use for entertainment would decrease the time on education, resulting in low academic performance [44]. Although the associations of SNS use patterns with SNS addiction are little known, the associations between problematic uses of smartphones or the internet, and addictive behaviors have been identified in the literature. Excessive use or habitual checking of the smartphone can lead to smartphone addiction [53-55]. Similarly, long durations of time spent on the internet are a significant indicator of internet addiction [56]. Moreover, previous studies have reported that young adults with internet and smartphone addiction mainly spend time on the use of SNS websites and applications [57,58]. Given that SNSs can be used over the internet through smartphones and other electronic devices, and that internet and smartphone addiction results in similar user behaviors, our findings on SNS addiction can be comparable to studies investigating internet use on smartphones or internet addiction. Moreover, shorter durations from waking to first SNS use and longer durations before sleeping are strong indicators of SNS addiction [53,59]. These findings from previous studies resembled our findings regarding the association of excessive SNS use patterns with SNS addiction. We also further found the effect of the SNS use patterns on SNS addiction, namely that excessive SNS use can cause SNS addiction.

Although excessive SNS use can cause SNS addiction behaviors, young people are poor at recognizing their own SNS use as excessive. In 1 study that surveyed the effects of Facebook use on the social life and behaviors of 1000 university students in Pakistan, nearly 70% of those who showed addictive Facebook use did not discern themselves as having Facebook addiction [60]. Similarly, another study that collected SNS information on 804 adolescents in India revealed that only 4% perceived their SNS use to be highly problematic [61]. Opportunities exist for health professionals to guide young people to recognize their own excessive SNS use that can cause SNS addictive behaviors. Practical screening for SNS addiction can be assessed as a daily routine in the community when dealing with young people who show problematic and excessive SNS use. This may be the first

step in recognizing potential cases before they develop further into SNS addiction. Concerning previous treatment for internet addiction [62], interventions conducted through a cognitive behavioral approach may also be helpful for users with SNS addiction in order to reconstruct their thoughts and behaviors and mitigate mental distress stemming from SNS addiction.

Impact of SNS Addiction on Mental Health Status

We found that SNS addition significantly influences mental health status (ie, anxiety and depressive symptoms), while the effect of SNS use patterns on mental health status is not significant. It should be noted that the difference between SNS use and SNS addiction is that the latter reflects more problematic behaviors and can lead to more severe outcomes that are associated with SNSs [7]. General SNS use may not be a potential factor; instead, it is the excessive use of SNSs that influences addictive behavior and mental distress [63]. Similar to our results, previous studies have reported that SNS addiction is associated with negative mental health outcomes, including depression, anxiety, and affective disorder [26,64,65]. However, these studies have mainly focused on the bivariate association between SNS addiction and mental health, while the effect of SNS addiction on mental health remains uncertain; that is, whether mental problems lead to SNS addiction or vice versa.

Furthermore, we identified that SNS addiction plays a mediating role in the pathway from SNS use to mental health status. This finding resonates with 1 study that identified that Instagram (an SNS) use predicts depression [25]. However, the mediating role of SNS addiction was not significant in that study, in contrast with our findings. SNS use in that study was only measured by 1 item (ie, the time spent on Instagram) [25], while we included diverse SNS use patterns (ie, 4 durations and 1 frequency measurement of SNS use). We also included all types of SNS use, but the aforementioned study only included Instagram use, although the use of different SNSs would be differently associated with SNS addiction and depression [25]. Therefore, the findings from our study offer more holistic and integrative viewpoints on the pathway from SNS use to mental health status via SNS addiction.

Of note, SNS use in moderation can positively affect users' mental health status. SNS use can reinforce a user's online relationships and solidify their offline connections, and this may, in turn, reduce their negative feelings and emotions (eg, anxiety, depression), positively influencing their mental health status [66,67]. Unfortunately, when SNS use becomes excessive, it conversely has an adverse effect on users. Although excessive SNS use is a popular concept in SNS studies and is known as a predictor of SNS addiction [68,69], there is ambiguity in the cut-off between non-excessive and excessive use, making excessive SNS use difficult to define. Therefore, SNS addiction would be more suitable than excessive SNS use to objectively determine that a user's SNS use is problematic and to understand its influence on health outcomes [61,69]. Moreover, the sense of belongingness (feeling of being accepted in groups) is high among young people, particularly from East Asian countries that have collectivism-dominant societies. The sense of belongingness is positively associated with social and psychological functioning [70,71]. The characteristics of young people and the social environment in HK (ie, a high sense of belongingness that we measured in this study using social well-being, loneliness, and isolation scales) likely mediate SNS use, SNS addiction, and mental health status.

As an indispensable element of today's leisure culture, absolute abstinence from accessing SNSs is not an appropriate treatment for SNS addiction and mental distress. Restricting the excessive use of SNSs, addressing the importance of SNS addiction control to improve psychosocial and mental health status and preventing relapse by encouraging self-reflection on SNS use may be possible solutions for designing educational programs. Particularly, during the COVID-19 pandemic, which hinders face-to-face intervention delivery, internet-delivered interventions (eg, internet-delivered cognitive behavioral therapy) can be also considered to address SNS addiction. Future studies on determining the appropriate duration of SNS use with findings transferable to practice guidelines would be needed.

Limitations

This study had some limitations. Cross-sectional data in this study could not provide the causality between SNS addiction and mental health status. Discussions of the magnitude of the relationship between 2 elements (SNS addiction and mental health status) were inconclusive, with different opinions that this relationship can be bidirectional [72,73]. Therefore, more studies with longitudinal designs are needed to explore the direction and mechanism of causality for the relationship between SNS addiction and mental health status in the future. Additionally, as most of our participants were ethnic Chinese in HK, replication of our study for young adults in other sociocultural settings is recommended to reveal a more comprehensive relationship that was discovered through this study. Lastly, the participants responded to an online survey. Self-reported data may result in recall bias.

Conclusion

This study provided novel information about the patterns of SNS use and its association with SNS addiction among university students. Findings from SEM also addressed that there is a significant mediating effect of SNS addiction between SNS use and mental health status, including anxiety and depressive symptoms. Further studies are suggested to demonstrate causal relationships with longitudinal data. This study helps to provide preliminary solutions for reducing SNS addiction and mental problems by conducting interventions using cognitive-behavioral approaches.

Acknowledgments

The authors would like to thank all students who assisted with the study design and participant recruitment for this study: Mr Kwong Ming Hong Bernard, Mr Lee Tsan Ho, Ms Chan Hue Yuet Kalfanie, Ms To Yi Yan Jenny, Mr Or Ego, Ms Cheung Tsz

Ching, Ms Ng Yeuk Tung, Ms Leung Ka Yan, Ms Mak Hoi Yin Tiffany, and Mr Man Anthony Gar-Ta from the University of Hong Kong.

Authors' Contributions

TW and JJL undertook the data analysis. TW drafted the manuscript. JJL conceived the study design. ACYL designed the survey and helped recruit the participants. JYHW, MPW, and SSK supported the study supervision. JJL, JYHW, MPW, and SSK revised the manuscript. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Griffiths MD. Social networking addiction: emerging themes and issues. J Addict Res Ther 2013;04(05):1000e118. [doi: 10.4172/2155-6105.1000e118]
2. Kuss D, Griffiths M. Social networking sites and addiction: ten lessons learned. Int J Environ Res Public Health 2017 Mar 17;14(3):311 [FREE Full text] [doi: 10.3390/ijerph14030311] [Medline: 28304359]
3. Phua J, Jin SV, Kim JJ. Uses and gratifications of social networking sites for bridging and bonding social capital: a comparison of Facebook, Twitter, Instagram, and Snapchat. Comput Hum Behav 2017 Jul;72:115-122. [doi: 10.1016/j.chb.2017.02.041]
4. Legislative Council Secretariat. Social Media Usage in Hong Kong. 2019 Dec 12. URL: https://www.legco.gov.hk/research-publications/english/1920issh15-social-media-usage-in-hong-kong-20191212-e.pdf [accessed 2020-07-08]
5. Census Statistics Department. Thematic Household Survey Report No 69: Personal Computer and Internet Penetration. 2019. URL: https://www.statistics.gov.hk/pub/B11302692020XXXXB0100.pdf [accessed 2021-05-06]
6. Census and Statistics Department. Thematic Household Survey Report No. 64: Personal Computer and Internet Penetration. 2017. URL: https://www.statistics.gov.hk/pub/B11302642018XXXXB0100.pdf [accessed 2021-10-21]
7. Zheng X, Lee MK. Excessive use of mobile social networking sites: negative consequences on individuals. Comput Hum Behav 2016 Dec;65:65-76. [doi: 10.1016/j.chb.2016.08.011]
8. Echeburúa E, De Corral P. Adicción a las nuevas tecnologías y a las redes sociales en jóvenes: un nuevo reto. Adicciones 2010 Jun 01;22(2):91. [doi: 10.20882/adicciones.196]
9. Barbera L, Paglia L, Valsavoia R. Social network and addiction. Stud Health Technol Inform 2009;144:33-36. [doi: 10.3233/978-1-60750-017-9-33]
10. AlAbdulwahab SS, Kachanathu SJ, AlMotairi MS. Smartphone use addiction can cause neck disability. Musculoskelet Care 2017 Mar 19;15(1):10-12. [doi: 10.1002/msc.1170] [Medline: 28105706]
11. Wang J, Jackson LA, Gaskin J, Wang H. The effects of social networking site (SNS) use on college students' friendship and well-being. Comput Hum Behav 2014 Aug;37:229-236. [doi: 10.1016/j.chb.2014.04.051]
12. Liu D, Kirschner PA, Karpinski AC. A meta-analysis of the relationship of academic performance and social network site use among adolescents and young adults. Comput Hum Behav 2017 Dec;77:148-157. [doi: 10.1016/j.chb.2017.08.039]
13. Huang C. Social network site use and academic achievement: a meta-analysis. Comput Educ 2018 Apr;119:76-83. [doi: 10.1016/j.compedu.2017.12.010]
14. Baek YM, Bae Y, Jang H. Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):512-517. [doi: 10.1089/cyber.2012.0510] [Medline: 23697533]
15. Pantic I. Online social networking and mental health. Cyberpsychol Behav Soc Netw 2014 Oct;17(10):652-657 [FREE Full text] [doi: 10.1089/cyber.2014.0070] [Medline: 25192305]
16. 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]
17. Lin R, Utz S. The emotional responses of browsing Facebook: happiness, envy, and the role of tie strength. Comput Hum Behav 2015 Nov;52:29-38 [FREE Full text] [doi: 10.1016/j.chb.2015.04.064] [Medline: 26877584]
18. 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]
19. Salehan M, Negahban A. Social networking on smartphones: when mobile phones become addictive. Comput Hum Behav 2013 Nov;29(6):2632-2639. [doi: 10.1016/j.chb.2013.07.003]
20. Müller K, Dreier M, Beutel M, Duven E, Giralt S, Wölfling K. A hidden type of internet addiction? Intense and addictive use of social networking sites in adolescents. Comput Hum Behav 2016 Feb;55:172-177. [doi: 10.1016/j.chb.2015.09.007]
21. Andreassen CS. Online social network site addiction: a comprehensive review. Curr Addict Rep 2015 Jun 01;2(2):175-184. [doi: 10.1007/s40429-015-0056-9]
22. Griffiths M, Kuss D, Demetrovics Z. Social networking addiction: an overview of preliminary findings. In: Rosenberg KP, Feder LC, editors. Behavioral Addictions: Criteria, Evidence, and Treatment. London: Academic Press; 2014:119-141.

23. Andreassen CS, Griffiths MD, Gjertsen SR, Krossbakken E, Kvam S, Pallesen S. The relationships between behavioral addictions and the five-factor model of personality. J Behav Addict 2013 Jun;2(2):90-99. [doi: 10.1556/JBA.2.2013.003] [Medline: 26165928]
24. Larose R, Mastro D, Eastin MS. Understanding internet usage. Soc Sci Comput Rev 2016 Aug 19;19(4):395-413. [doi: 10.1177/089443930101900401]
25. Donnelly E. Depression among users of social networking sites (SNSs): the role of SNS addiction and increased usage. J Add Pre Med 2017 May 24;02(01):107. [doi: 10.19104/japm.2016.107]
26. Tang CS, Koh YYW. Online social networking addiction among college students in Singapore: comorbidity with behavioral addiction and affective disorder. Asian J Psychiatr 2017 Feb;25:175-178. [doi: 10.1016/j.ajp.2016.10.027] [Medline: 28262144]
27. Jackson LA, Wang J. Cultural differences in social networking site use: a comparative study of China and the United States. Comput Hum Behav 2013 May;29(3):910-921. [doi: 10.1016/j.chb.2012.11.024]
28. Vizzotto A, Oliveira A, Elkis H, Cordeiro Q, Buchain P. Psychosocial characteristics. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York: Springer; 2013:1578-1580.
29. Cho J. Roles of smartphone app use in improving social capital and reducing social isolation. Cyberpsychol Behav Soc Netw 2015 Jun;18(6):350-355. [doi: 10.1089/cyber.2014.0657] [Medline: 26075923]
30. Stansfeld S, Rasul F. Psychosocial factors, depression and illness. In: Steptoe A, editor. Depression and Physical Illness. Cambridge: Cambridge University Press; 2007:19-49.
31. Yang C. Instagram use, loneliness, and social comparison orientation: interact and browse on social media, but don't compare. Cyberpsychol Behav Soc Netw 2016 Dec;19(12):703-708. [doi: 10.1089/cyber.2016.0201] [Medline: 27855266]
32. Koc M, Gulyagci S. Facebook addiction among Turkish college students: the role of psychological health, demographic, and usage characteristics. Cyberpsychol Behav Soc Netw 2013 Apr;16(4):279-284. [doi: 10.1089/cyber.2012.0249] [Medline: 23286695]
33. Wang P, Wang X, Wu Y, Xie X, Wang X, Zhao F, et al. Social networking sites addiction and adolescent depression: a moderated mediation model of rumination and self-esteem. Pers Individ Differ 2018 Jun;127:162-167. [doi: 10.1016/j.paid.2018.02.008]
34. Pontes HM. Investigating the differential effects of social networking site addiction and Internet gaming disorder on psychological health. J Behav Addict 2017 Dec 01;6(4):601-610 [FREE Full text] [doi: 10.1556/2006.6.2017.075] [Medline: 29130329]
35. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook addiction scale. Psychol Rep 2012 Apr 01;110(2):501-517. [doi: 10.2466/02.09.18.pr0.110.2.501-517]
36. Andreassen CS, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
37. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ–4. Psychosomatics 2009 Nov;50(6):613-621. [doi: 10.1016/s0033-3182(09)70864-3]
38. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2009 May 28;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]
39. Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist 2006 Aug 01;46(4):503-513. [doi: 10.1093/geront/46.4.503] [Medline: 16921004]
40. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004 Aug 19;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
41. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
42. Yu CY, Muthén B. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. In: Doctoral Dissertation. Los Angeles, CA: University of California; 2002:11-17.
43. Wang C, Ho RT, Chan CL, Tse S. Exploring personality characteristics of Chinese adolescents with internet-related addictive behaviors: trait differences for gaming addiction and social networking addiction. Addict Behav 2015 Mar;42:32-35. [doi: 10.1016/j.addbeh.2014.10.039] [Medline: 25462651]
44. Kircaburun K. Effects of gender and personality differences on Twitter addiction among Turkish undergraduates. J Educ Pract 2016;7(24):33-42 [FREE Full text]
45. Mazman S, Usluel Y. Gender differences in using social networks. Turkish Online J Educ Technol 2011;10(2):133-139 [FREE Full text]
46. Błachnio A, Przepiórka A, Pantic I. Internet use, Facebook intrusion, and depression: results of a cross-sectional study. Eur Psychiatry 2015 Sep;30(6):681-684. [doi: 10.1016/j.eurpsy.2015.04.002] [Medline: 25963476]

47. Fu G, Xu F, Cameron CA, Leyman G, Lee K. Cross-cultural differences in children's choices, categorizations, and evaluations of truths and lies. Dev Psychol 2007 Mar;43(2):278-293 [FREE Full text] [doi: 10.1037/0012-1649.43.2.278] [Medline: 17352539]
48. Lee Y, Chang C, Lin Y, Cheng Z. The dark side of smartphone usage: psychological traits, compulsive behavior and technostress. Comput Hum Behav 2014 Feb;31:373-383. [doi: 10.1016/j.chb.2013.10.047]
49. Glass R, Li S, Pan R. Personality, problematic social network use and academic performance in China. J Comput Inf Syst 2015 Dec 10;54(4):88-96. [doi: 10.1080/08874417.2014.11645726]
50. Azizi SM, Soroush A, Khatony A. The relationship between social networking addiction and academic performance in Iranian students of medical sciences: a cross-sectional study. BMC Psychol 2019 May 03;7(1):28 [FREE Full text] [doi: 10.1186/s40359-019-0305-0] [Medline: 31053171]
51. AlBarashdi HS. Social networking (SNS) addiction among university students: a literature review and research directions. JESBS 2020 Feb 18:11-23. [doi: 10.9734/jesbs/2020/v33i130191]
52. Maqableh M, Rajab L, Quteshat W, Masa'deh RMT, Khatib T, Karajeh H. The impact of social media networks websites usage on students' academic performance. Commun Netw 2015;07(04):159-171. [doi: 10.4236/cn.2015.74015]
53. 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]
54. Cha S, Seo B. Smartphone use and smartphone addiction in middle school students in Korea: prevalence, social networking service, and game use. Health Psychol Open 2018 Feb;5(1):2055102918755046 [FREE Full text] [doi: 10.1177/2055102918755046] [Medline: 29435355]
55. Matar Boumosleh J, Jaalouk D. Depression, anxiety, and smartphone addiction in university students: a cross sectional study. PLoS One 2017 Aug 4;12(8):e0182239 [FREE Full text] [doi: 10.1371/journal.pone.0182239] [Medline: 28777828]
56. Weinstein A, Lejoyeux M. Internet addiction or excessive internet use. Am J Drug Alcohol Abuse 2010 Sep 15;36(5):277-283. [doi: 10.3109/00952990.2010.491880] [Medline: 20545603]
57. Lopez-Fernandez O, Honrubia-Serrano L, Freixa-Blanxart M, Gibson W. Prevalence of problematic mobile phone use in British adolescents. Cyberpsychol Behav Soc Netw 2014 Feb;17(2):91-98. [doi: 10.1089/cyber.2012.0260] [Medline: 23981147]
58. Li S, Chung T. Internet function and internet addictive behavior. Comput Hum Behav 2006 Nov;22(6):1067-1071. [doi: 10.1016/j.chb.2004.03.030]
59. Lee JJ, Wang MP, Luk TT, Guo N, Chan SS, Lam TH. Associations of electronic device use before and after sleep with psychological distress among Chinese adults in Hong Kong: cross-sectional study. JMIR Ment Health 2020 Jun 11;7(6):e15403 [FREE Full text] [doi: 10.2196/15403] [Medline: 32525489]
60. Farooqi H, Patel H, Aslam H, Ansari I, Khan M, Iqbal N, et al. Effect of Facebook on the life of medical university students. Int Arch Med 2013 Oct 17;6(1):40 [FREE Full text] [doi: 10.1186/1755-7682-6-40] [Medline: 24134850]
61. Dhir A, Kaur P, Chen S, Lonka K. Comput Hum Behav 2016 Jun;59:420-430. [doi: 10.1016/j.chb.2016.02.040]
62. Young KS. Internet addiction: evaluation and treatment. BMJ 1999 Oct 01;319(Suppl S4):9910351. [doi: 10.1136/sbmj.9910351]
63. Kuss DJ, Griffiths MD. Online social networking and addiction: a review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]
64. Frost RL, Rickwood DJ. A systematic review of the mental health outcomes associated with Facebook use. Comput Hum Behav 2017 Nov;76:576-600. [doi: 10.1016/j.chb.2017.08.001]
65. Hong F, Huang D, Lin H, Chiu S. Analysis of the psychological traits, Facebook usage, and Facebook addiction model of Taiwanese university students. Telemat Inform 2014 Nov;31(4):597-606. [doi: 10.1016/j.tele.2014.01.001]
66. Grieve R, Indian M, Witteveen K, Anne Tolan G, Marrington J. Face-to-face or Facebook: can social connectedness be derived online? Comput Hum Behav 2013 May;29(3):604-609. [doi: 10.1016/j.chb.2012.11.017]
67. Frison E, Eggermont S. Exploring the relationships between different types of Facebook use, perceived online social support, and adolescents depressed mood. Soc Sci Comput Rev 2015 Jan 21;34(2):153-171. [doi: 10.1177/0894439314567449]
68. Wohn DY, LaRose R. Effects of loneliness and differential usage of Facebook on college adjustment of first-year students. Comput Educ 2014 Jul;76:158-167. [doi: 10.1016/j.compedu.2014.03.018]
69. Muench F, Hayes M, Kuerbis A, Shao S. The independent relationship between trouble controlling Facebook use, time spent on the site and distress. J Behav Addict 2015 Sep;4(3):163-169 [FREE Full text] [doi: 10.1556/2006.4.2015.013] [Medline: 26551906]
70. Hagerty BM, Williams RA, Coyne JC, Early MR. Sense of belonging and indicators of social and psychological functioning. Arch Psychiatr Nurs 1996 Aug;10(4):235-244. [doi: 10.1016/s0883-9417(96)80029-x]
71. Ahn MY, Davis HH. Four domains of students' sense of belonging to university. Stud High Educ 2019 Jan 09;45(3):622-634. [doi: 10.1080/03075079.2018.1564902]
72. Clayton RB, Osborne RE, Miller BK, Oberle CD. Loneliness, anxiousness, and substance use as predictors of Facebook use. Comput Hum Behav 2013 May;29(3):687-693. [doi: 10.1016/j.chb.2012.12.002]

73. Sampasa-Kanyinga H, Lewis RF. Frequent use of social networking sites is associated with poor psychological functioning among children and adolescents. Cyberpsychol Behav Soc Netw 2015 Jul;18(7):380-385. [doi: 10.1089/cyber.2015.0055] [Medline: 26167836]

Abbreviations

aβ: adjusted beta
BC: bias-corrected
CFI: comparative fit index
DSM-5: Diagnostic and Statistical Manual 5th Edition
HK: Hong Kong
ICD-11: International Classification of Diseases 11th Revision
RMSEA: root-mean-square error of approximation
SEM: structural equation modeling
SNS: social networking service
SRMR: standardized root-mean-square residual
TLI: Tucker-Lewis index
UGC: University Grants Committee

Edited by R Kukafka; submitted 28.12.20; peer-reviewed by H Norman, B Parker; comments to author 08.03.21; revised version received 26.04.21; accepted 19.10.21; published 08.12.21.

Please cite as:
Wang T, Wong JYH, Wang MP, Li ACY, Kim SS, Lee JJ
Effects of Social Networking Service (SNS) Addiction on Mental Health Status in Chinese University Students: Structural Equation Modeling Approach Using a Cross-sectional Online Survey
J Med Internet Res 2021;23(12):e26733
URL: https://www.jmir.org/2021/12/e26733
doi:10.2196/26733
PMID:34889760

Original Paper

Frequency of Online Health Information Seeking and Types of Information Sought Among the General Chinese Population: Cross-sectional Study

Zihui Xiong[1,2], MA; Liang Zhang[3], MA; Zhong Li[4], PhD; Wanchun Xu[1,2], MA; Yan Zhang[1,2], PhD; Ting Ye[1,2], PhD

[1]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Research Center for Rural Health Service, Key Research Institute of Humanities & Social Sciences of Hubei Provincial Department of Education, Wuhan, China

[3]School of Political Science and Public Administration, Wuhan University, Wuhan, China

[4]School of Health Policy and Management, Nanjing Medical University, Nanjing, China

Corresponding Author:
Ting Ye, PhD
School of Medicine and Health Management
Tongji Medical College
Huazhong University of Science and Technology
No 13 Hangkong Rd
Qiaokou District
Wuhan, 430030
China
Phone: 86 134 7621 5639
Email: yeting@hust.edu.cn

Abstract

Background: The internet is one of the most popular health information resources, and the Chinese constitute one-fifth of the online users worldwide. As internet penetration continues to rise, more details on the Chinese population seeking online health information need to be known based on the current literature.

Objective: This study aims to explore the sociodemographic differences in online health information seeking (OHIS), including the frequency of OHIS and the types of online health information sought among the general Chinese population in mainland China.

Methods: A cross-sectional study for assessing the residents' health care needs with self-administered questionnaires was implemented in 4 counties and districts in China from July 2018 to August 2018. Pearson's chi-square test was used to identify the sociodemographic differences between infrequent and frequent online health information seekers. We also performed binary logistic regression for the 4 types of online health information as the dependent variables and the sociodemographic factors as the independent variables.

Results: Compared with infrequent online health information seekers, frequent seekers were more likely to be female (infrequent: 1654/3318; 49.85%; frequent: 1015/1831, 55.43%), older (over 60 years old; infrequent: 454/3318, 13.68%; frequent: 282/1831, 15.40%), married (infrequent: 2649/3318, 79.84%; frequent: 1537/1831, 83.94%), and better educated (bachelor's or above; infrequent: 834/3318, 25.14%; frequent: 566/1831, 30.91%). They were also more likely to earn a higher income (over RMB ¥50k [RMB ¥1=US $0.15641]; infrequent: 1139/3318, 34.33%; frequent: 710/1831, 34.78%), have commercial health insurance (infrequent: 628/3318, 18.93%; frequent: 470/1831, 25.67%), and have reported illness in the past 12 months (infrequent: 659/3318, 19.86%; frequent: 415/1831, 22.67%). Among the 4 health information types, health science popularization was the most searched for information by Chinese online health information seekers (3654/5149, 70.79%), followed by healthy behaviors (3567/5149, 69.28%), traditional Chinese medicine (1931/5149, 37.50%), and medical concerns (1703/5149, 33.07%). The binary logistic regression models showed that males were less likely to seek information on healthy behaviors (adjusted odds ratio [AOR] 0.69, 95% CI 0.61-0.78) and traditional Chinese medicine (AOR 0.64, 95% CI 0.57-0.73), and respondents who had at least 1 chronic disease were more likely to seek information on medical concerns (AOR 1.27, 95% CI 1.07-1.51) and traditional Chinese medicine (AOR 1.26, 95% CI 1.06-1.49).

Conclusions: Sociodemographic factors were associated with the frequency of OHIS and types of information sought among the general Chinese population. The results remind providers of online health information to consider the needs of specific population groups when tailoring and presenting health information to the target population.

(J Med Internet Res 2021;23(12):e30855) doi:10.2196/30855

KEYWORDS

online health information seeking; sociodemographic factors; information types; Chinese population; information seeking behavior; demography; China; online health information

Introduction

Given its rapid development, the internet has become an essential source of health information worldwide. According to Internet World Stats, the number of internet users has reached 5 billion, accounting for more than 60% of the world population [1]. The internet has the highest usage rate among health information sources [2-4]. Although certain groups still rely on traditional media for health information [5], the internet is increasingly growing in popularity as a source of health information.

The advantages of online health information seeking (OHIS) manifest in various aspects. Turning to the internet before seeking a professional diagnosis has the potential to improve the relationship between patients and their physicians, and patients are more inclined to trust in their physicians' advice by discussing their findings online with these professionals [6-8]. The likelihood of individuals changing their health behavior rises as the frequency of using the internet to retrieve health information increases [9]. For people who live with chronic diseases, the internet may improve their skills in managing their condition [10,11]. However, seeking health information online can be a mixed blessing. Studies have shown that online health information can be inadequate and inaccurate [12-14]. The information may be of questionable quality, leading to hasty, ill-informed, and dangerous health choices [15]. Compared with that available in Western countries, the low-quality online health information seems to be particularly serious in Asian countries [16].

Previous studies have explored how sociodemographic factors influence OHIS behavior. The frequency of OHIS is higher among individuals who are female, younger, better educated, or have a higher household income [9,17,18]. A study that analyzed 2 Pew data sets collected in 2002 and 2012 showed that the sociodemographic factors of gender and education strongly predicted OHIS in both years; meanwhile, age and income were significant predictors in 2012 but not in 2002 [19]. Internet access and OHIS behavior have been shown to vary by race and ethnicity [20-24]. Other factors, such as health status [17] and employment status [24], were also found to be statistically significant in OHIS. A meta-analysis on OHIS revealed that age, gender, education, income, race, and experience were social determinants of OHIS [25]. In summary, most related studies indicate that the crucial characteristics of online health information seekers are age, level of education, income, and ethnicity [26]. Several studies have consistently focused on the predictors of OHIS; however, most of them aimed at identifying determinants that discriminate between online health information seekers and nonseekers [4,20,24,27-29], and elements associated with OHIS among those who have used the internet for health information should also be examined.

Seekers of online health information seek various health topics. Some studies reported information about disease to be the most common topic sought by online health information seekers [19,30]. Other frequently searched topics include symptoms, medication, treatment, exercise and fitness, and nutrition or diet [8,10,30,31]. However, there is no official or agreed upon classification for these common health topics. By reviewing previous studies, these topics can be mainly grouped into 2 types: (1) healthy behaviors, including information related to nutrition or diet, exercise, and body maintenance; (2) and medical concerns, including information related to disease, medications, and treatments [19,32]. Of these 2 types, information on medical concerns is more likely to be sought by people with more health risks [8,31,32]. For instance, compared with the population with no chronic disease, people with chronic conditions are more likely to search about medicines [10]. Overall, evidence shows that medical information seems to be the most popular health information on the internet. However, different population groups have different preferences for health information, and how individuals who seek this information may vary by their characteristics remains obscure.

According to the 47th Statistical Report on China's internet Development [33], as of December 2020, 9.89 million Chinese had access to the internet, and the penetration rate in China reached 70.4%, a rate higher than the average level in Asia (62.6%) [1]. The Chinese population constitutes one-fifth of the world's internet users, and easy access to specialized information, such as medical and health care information, accounts for 29.8% of the factors facilitating China's nonnetizens' access to the internet [33]. However, academic research on OHIS has mostly collected evidence from Western countries, especially the United States [5,10,21], or has focused on certain groups, such as women [23,34], adolescents [28,35], and patients with chronic disease [10,11]. Studies on the general Chinese population, particularly the Mainland population, are still limited.

According to the literature review, OHIS has not only positive but also negative influences on individuals. With the rise of the internet penetration rate, an increasing number of people will have access to online health information. New health information is difficult to effectively categorize when individuals are overwhelmed by diverse or mixed health information, and an overabundance of health information on the internet can make individuals feel confused and become more anxious about

their conditions [36,37]. To ensure that online health information can be effectively used and to understand how individuals can best avoid the negative influence of OHIS, it is crucial to understand the characteristics and preferences of online health information seekers so that practical information can be tailored and presented to the target population. As a result, we aimed to answer 3 questions: (1) Which groups of people are frequent online health information seekers? (2) Which types of online health information do people usually search for? (3) What are the preferences for types of online health information among different population groups? Given that sociodemographic factors have provided an increasingly significant explanation of the variance in OHIS [19], our study aimed to explore sociodemographic differences in OHIS behavior among the general Chinese population, including frequency of OHIS and types of health information sought.

This study used the data of a household survey conducted in 4 counties and districts in Mainland China, which may additionally contribute to the current research. As inaccurate information and the inability to find information may discourage people from retrieving professional information [38], our findings may offer additional knowledge to online health information providers so that appropriate health information content can be tailored and presented. Therefore, we believe this study can significantly contribute to online health information management and promotion in China.

Methods

Study Design

From July 2018 to August 2018, the Research Center for Rural Health Services carried out a study to assess residents' health care needs. Based on the assessment of socioeconomic development and geographic distribution and suggestions from experts [39], 4 counties and districts across China were purposely selected as the survey areas, including 2 districts in urban China (Xiling in central China, Futian in eastern China) and 2 counties in rural China (Dangyang in central China, Sinan in western China) [40].

A multistage stratified random sampling was implemented to select representative samples of the districts' and counties' populations. In the sample size calculation, the design effect was set at 2.5, and the allowable error at the significance level of .05; the chronic disease prevalence was 21.338% according to the 2013 National Health Service Survey [41], and thus the minimum sample size in each survey area was calculated to be 3584 individuals. Given that a family in China has an average of 2.9 members [41], at least 1235 families in every study site needed to be investigated. In the 2 counties (Dangyang and Sinan), 5 towns in each county were randomly selected according to the distance to the county hospitals. Meanwhile, 6 villages in each town were randomly selected according to the distance to the township hospitals. In the 2 districts (Xiling and Futian), 5 streets in each district were randomly selected according to the distance to the medical centers. To ensure that at least 248 families in each town or street were effectively interviewed, 1 other town or street in each county or district was selected as a backup in case of refusal, closed household, or removal of defective questionnaires. We resampled and surveyed until the final sample size met the estimated requirement. The surveyed families were systematically sampled in the rosters of residents from the village or street councils. Under the guidance of well-trained interviewers, face-to-face interviews were conducted, and self-administrated questionnaires were used on all the members of the sampled families. Finally, data from 15,126 respondents in 5547 families were collected in total: 3310 individuals in 1360 families in 30 villages in Dangyang, 3983 individuals in 1355 families in 30 villages in Sinan, 4386 individuals in 1513 families in 6 streets in Futian, and 3447 individuals in 1319 families in 5 streets in Xiling. A total of 12,646 individuals above 15 years of age needed to answer the questions related to OHIS, and 5354 respondents (5354/12,646, 42.34%) reported using the internet to seek health information. Among the online health information seekers, those who failed to provide information for all variables queried in this study were excluded, and 5149 individuals were finally included in the analysis.

Ethical approval was obtained from the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (#IORG0003571).

Variables

Sociodemographic Factors

The survey questionnaires were divided into several parts, including basic family information, individual information, health status, health care use, and health promotion need. Sociodemographic factors, including gender, age, marital status, area of residence, education level, annual household income per capita, employment, health insurance, having at least 1 chronic disease, and reported illness in the past 12 months, were considered to be independent variables. The coding of independent variables is shown in Table 1.

Table 1. Coding of independent variables.

Variables	Assignment
Gender	Female=0; male=1
Age (years)	15-45=1; 46-60=2; >60=3
Marital status[a]	Single=0; married=1
Area	Rural=0; urban=1
Education level	Junior high school or below=1; senior high school =2; 3-year college=3; bachelor's or above=4
Annual household income per capita (RMB ¥[b])	<10k=1; 10k-30k=2; 30k-50k=3; >50k=4
Employment[c]	Unemployed=0; employed=1
Types of primary health insurance	Urban employee basic medical insurance=1; urban-rural resident basic medical insurance=2; others[d]=3
Having commercial health insurance	No=0; yes=1
Having at least 1 chronic disease	No=0; yes=1
Reported illness in the past 12 months	No=0; yes=1

[a]Marital status was divided into 2 categories: single (unmarried, divorced, or widowed) and married (married).

[b]RMB ¥1=US $0.15641.

[c]Employment was recoded as 2 groups: unemployed (unemployed, students, and retired) and employed (employed).

[d]Others included no primary health insurance and health insurance outside mainland China.

Frequency of OHIS

Most frequency response scales use subjective terms as scale labels, such as the 3-point frequency scale (usually, sometimes or partially, never) and 4-point frequency scale (never, seldom/hardly ever, sometimes, often) [9,42,43]. In this study, the measurement of the frequency of OHIS was adapted from previous studies, which were based on respondents' subjective ratings. Respondents were asked to rate the frequency of OHIS on a 3-point frequency scale (frequent, infrequent, never), and those who selected "Frequent" and "Infrequent" were classified as frequent and infrequent online health information seekers, respectively. Given that our study participants were those who had used online health information, respondents who selected "Never" were excluded.

Types of Online Health Information

In the original survey, respondents were asked whether they had searched for the following health-related contents: sports and fitness, health science popularization, traditional Chinese medicine (TCM), women's health, nutrition and diet, medication guidance, disease consultation, health status monitoring, and others. In this study, 6 items were selected and then grouped into 4 types to be analyzed: (1) healthy behaviors, including sport and fitness, and nutrition and diet; (2) medical concerns, including medication guidance and disease consulting; (3) TCM, which may particularly be sought by Chinese users [31,44]; and (4) health science popularization, an emerging type of health information that aims at improving people's health literacy, which the National Health Commission of China has strongly supported in recent years. According to the National Health Commission of China, information on health science popularization mainly includes scientific knowledge about health technologies, ideas, methods, and skills, which is particularly disseminated in a way that is easy to understand, accept, and participate in [45]. Of the 4 types of online health information, the first 2 contain the common health topics frequently sought by the general population worldwide, and the analysis of the latter 2 may be helpful to better understanding the OHIS behavior of the general population in the setting of Mainland China. Each type as a dependent variable was dichotomized (yes or no).

Statistical Analysis

First, Pearson's chi-square test was used to compare sociodemographic differences between infrequent and frequent online health information seekers. Second, descriptive analysis was applied to show the prevalence of different types of online health information. Finally, binary logistic regression was conducted on each type of online health information to identify the sociodemographic differences in searches for health information types, and the adjusted odds ratio (AOR) and 95% CI were reported. The significance level was set at a *P* value of .05.

Results

Sociodemographic Information of the Respondents

Table 2 shows an overview of the main sociodemographic characteristics of the study population and differences in the frequency of OHIS. Online health information seekers comprised more females (2669/5149, 51.84), and most of them lived in urban China (3993/5149, 77.55%). Almost two-thirds (3318/5149, 64.44%) of respondents reported infrequently using the internet to seek health information. Among people who were frequent seekers of online health information, more than half (1015/1831, 55.43%) were female. The comparison between the 2 groups revealed that differences in terms of gender

(P<.001), age (P=.01), marital status (P<.001), education level (P<.001), annual household income per capita (P<.001), possession of commercial insurance (P<.001), and reported illness in the past 12 months (P=.02) were statistically significant. In contrast, area (P=.12), employment (P=.79), types of primary health insurance (P=.57), and having at least 1 chronic disease (P=.14) were not significantly different. Compared with infrequent online health information seekers, frequent seekers were more likely to be females (infrequent: 1654/3318, 49.85%; frequent: 1015/1831, 55.43%), were older (over 60 years old; infrequent: 454/3318, 13.68%; frequent: 282/1831, 15.40%), were more likely to be married (infrequent: 2649/3318, 79.84%; frequent: 1537/1831, 83.94%), were better educated (bachelor's or above; infrequent: 834/3318, 25.14%; frequent: 566/1831, 30.91%), earned a higher income (over RMB ¥50k [RMB ¥1=US $0.15641; frequent: 1139/3318, 34.33%; frequent: 710/1831, 34.78%), were more likely to have commercial health insurance (infrequent: 628/3318, 18.93%; frequent: 470/1831, 25.67%), and were more likely to have reported illness in the past 12 months (infrequent: 659/3318, 19.86%; frequent: 415/1831, 22.67%).

Table 2. Sociodemographic characteristics of the respondents by frequency of OHIS (online health information seeking).

Sociodemographic characteristics	Overall (N=5149), n (%)	Infrequent (N=3318), n (%)	Frequent (N=1831), n (%)	*P* value
Gender				<.001
Male	2480 (48.16)	1664 (50.15)	816 (44.57)	
Female	2669 (51.84)	1654 (49.85)	1015 (55.43)	
Age				.01
15-45 years	3015 (58.56)	1992 (60.04)	1023 (55.87)	
46-60 years	1398 (27.15)	872 (26.28)	526 (28.73)	
>60 years	736 (14.29)	454 (13.68)	282 (15.40)	
Marital status				<.001
Single	963 (18.70)	669 (20.16)	294 (16.06)	
Married	4186 (81.30)	2649 (79.84)	1537 (83.94)	
Area				0.12
Rural	1156 (22.45)	767 (23.12)	389 (21.25)	
Urban	3993 (77.55)	2551 (76.88)	1442 (78.75)	
Education level				<.001
Junior high school or below	1476 (28.67)	1029 (31.01)	447 (24.41)	
Senior high school	1547 (30.04)	996 (30.02)	551 (30.09)	
3-year college	726 (14.10)	459 (13.83)	267 (14.58)	
Bachelor's or above	1400 (27.19)	834 (25.14)	566 (30.91)	
Annual household income per capita[a]				<.001
<10k	578 (11.23)	406 (12.24)	172 (9.39)	
10k-30k	1600 (31.07)	1071 (32.28)	529 (28.89)	
30k-50k	1122 (21.79)	702 (21.16)	420 (22.94)	
>50k	1849 (35.91)	1139 (34.33)	710 (38.78)	
Employment				.79
Unemployed	1764 (34.26)	1141 (34.39)	623 (34.03)	
Employed	3385 (65.74)	2177 (65.61)	1208 (65.97)	
Types of primary health insurance				.57
UEBMI[b]	2331 (45.27)	1494 (45.03)	837 (45.71)	
URBMI[c]	2620 (50.88)	1702 (51.30)	918 (50.14)	
Other	198 (3.85)	122 (3.68)	76 (4.15)	
Having commercial health insurance				<.001
No	4051 (78.68)	2690 (81.07)	1361 (74.33)	
Yes	1098 (21.32)	628 (18.93)	470 (25.67)	
Having at least 1 chronic disease				.14
No	4154 (80.68)	2697 (66.40)	1457 (79.57)	
Yes	995 (19.32)	621 (57.19)	374 (20.43)	
Reported illness in the past 12 months				.02
No	4075 (79.14)	2659 (80.14)	1416 (77.33)	
Yes	1074 (20.86)	659 (19.86)	415 (22.67)	

[a]Income is reported in renminbi ¥. RMB ¥1=US $0.15641.

[b]UEBMI: urban employee basic medical insurance.

[c]URBMI: urban-rural resident basic medical insurance.

Types of Online Health Information Searched

Respondents reported 1 or more different types of health information they sought online. A majority of the respondents (3654/5149, 70.79%) had used the internet to seek health science popularization information, which was followed in popularity by healthy behaviors (3567/5149, 69.28%), TCM, and medical concerns (1703/5149, 33.07%). The prevalence of the searches for the 4 types of online health information is shown in Table 3.

Table 3. Prevalence of types of online health information sought by the study participants.

Types of online health information	Prevalence (N=5149), n (%)
Healthy behaviors	3567 (69.28)
Medical concerns	1703 (33.07)
Traditional Chinese medicine	1931 (37.50)
Health science popularization	3645 (70.79)

Sociodemographic Factors Associated With the Types of OHIS

The results of the binary logistic regression models illustrated sociodemographic differences in seeking the 4 types of online health information (Table 4). Being male was significantly associated with a lower possibility of seeking healthy behaviors (AOR 0.69, 95% CI 0.61-0.78) and TCM (AOR 0.64, 95% CI 0.57-0.73). Compared with individuals aged between 15 and 45 years, middle-aged individuals were more likely to seek TCM (AOR 1.86, 95% CI 1.61-2.15) and health science popularization (AOR 1.20, 95% CI 1.02-1.40), while older adults were more likely to seek information for medical concerns (AOR 1.40, 95% CI 1.12-1.76) and TCM (AOR 2.16, 95% CI 1.72-2.70). Compared with single respondents, married respondents were more likely to seek TCM (AOR 1.22, 95% CI 1.03-1.44) and health science popularization (AOR 1.20, 95% CI 1.02-1.41) but were less likely to seek healthy behaviors (AOR 0.67, 95% CI 0.56-0.80). Living in urban China was associated with increased odds of seeking TCM (AOR 1.23, 95% CI 1.01-1.52). Higher education attainment was correlated with a higher possibility of seeking healthy behaviors (3-year college: AOR 1.56, 95% CI 1.26-1.95; bachelor's or above: AOR 1.33, 95% CI 1.10-1.62), medical concerns (senior high school: AOR 1.19, 95% CI 1.01-1.39; bachelor's or above: AOR 1.46, 95% CI 1.20-1.77), and health science popularization (bachelor's or above: AOR 1.50, 95% CI 1.23-1.83). Lower annual household income per capita was associated with increased odds of seeking healthy behaviors (RMB ¥10k-30: AOR 1.21, 95% CI 1.02-1.44) but decreased odds of seeking TCM (<RMB ¥10k: AOR 0.70, 95% CI 0.54-0.91; RMB ¥10k-30k: AOR 0.76, 95% CI 0.64-0.89). Respondents with urban-rural resident basic medical insurance were less likely to seek health science popularization than those with urban employee basic medical insurance (AOR 0.86, 95% CI 0.74-1.00). Having commercial health insurance was associated with a higher possibility to seek healthy behaviors (AOR 1.55, 95% CI 1.32-1.83) and TCM (AOR 1.27, 95% CI 1.10-1.47) but a lower possibility to seek medical concerns (AOR 0.82, 95% CI 0.70-0.95) and health science popularization (AOR 0.83, 95% CI 0.71-0.96). Individuals who had at least 1 chronic disease were more likely to seek medical concerns (AOR 1.27, 95% CI 1.07-1.51) and TCM (AOR 1.26, 95% CI 1.06-1.49). Similarly, those who reported illness during the past 12 months were also more likely to seek medical concerns (AOR 1.41, 95% CI 1.21-1.65) and TCM (AOR 1.17, 95% CI 1.00-1.38).

Table 4. Binary logistic regression models for AORs and 95% CIs of reporting the search for each type of online health information.

Sociodemographic Characteristics	Healthy behaviors, AOR[a] (95% CI)	Medical concerns, AOR (95% CI)	Traditional Chinese medicine, AOR (95% CI)	Health science popularization, AOR (95% CI)
Gender				
Female (Ref[b])				
Male	0.69 (0.61-0.78)**	1.11 (0.98-1.25)	0.64 (0.57-0.73)**	1.13 (1.00-1.28)
Age				
15-45 years (Ref)				
46-60 years	1.07 (0.92-1.25)	0.93 (0.80-1.08)	1.86 (1.61-2.15)**	1.20 (1.02-1.40)*
>60 years	0.85 (0.67-1.07)	1.40 (1.12-1.76)**	2.16 (1.72-2.70)**	1.14 (0.90-1.46)
Marital status				
Single (Ref)				
Married	0.67 (0.56-0.80)**	1.17 (0.99-1.37)	1.22 (1.03-1.44)*	1.20 (1.02-1.41)*
Area				
Rural (Ref)				
Urban	0.97 (0.79-1.19)	0.95 (0.77-1.16)	1.23 (1.01-1.52)*	1.05 (0.85-1.29)
Education level				
Junior high school or below (Ref)				
Senior high school	1.16 (0.99-1.37)	1.19 (1.01-1.40)*	1.04 (0.89-1.22)	1.14 (0.97-1.34)
3-year college	1.56 (1.26-1.95)**	1.21 (0.98-1.50)	1.05 (0.85-1.29)	1.19 (0.96-1.48)
Bachelor's or above	1.33 (1.10-1.62)**	1.46 (1.20-1.77)**	0.89 (0.74-1.08)	1.50 (1.23-1.83)**
Annual household income per capita[c]				
>50k (Ref)				
<10k	0.94 (0.72-1.21)	0.84 (0.65-1.09)	0.70 (0.54-0.91)**	0.99 (0.76-1.28)
10k-30k	1.21 (1.02-1.44)*	1.03 (0.87-1.22)	0.76 (0.64-0.89)**	1.11 (0.94-1.32)
30k-50k	1.17 (0.98-1.38)	1.09 (0.92-1.28)	1.01 (0.86-1.19)	1.19 (1.00-1.41)
Employment				
Unemployed (Ref)				
Employed	0.88 (0.75-1.04)	1.10 (0.94-1.29)	1.12 (0.96-1.31)	0.92 (0.78-1.07)
Types of primary health insurance				
Urban employee basic medical insurance (Ref)				
URBMI[d]	0.96 (0.83-1.11)	1.05 (0.91-1.21)	1.06 (0.92-1.22)	0.86 (0.74-1.00)*
Other	0.93 (0.67-1.29)	1.12 (0.81-1.53)	0.79 (0.58-1.09)	0.69 (0.50-0.94)*
Having commercial health insurance				
No (Ref)				
Yes	1.55 (1.32-1.83)**	0.82 (0.70-0.95)*	1.27 (1.10-1.47)**	0.83 (0.71-0.96)*
Having at least 1 chronic disease				
No (Ref)				
Yes	1.02 (0.85-1.22)	1.27 (1.07-1.51)*	1.26 (1.06-1.49)**	0.96 (0.80-1.15)
Reported illness during the past 12 months				
No (Ref)				
Yes	0.88 (0.75-1.03)	1.41 (1.21-1.65)**	1.17 (1.00-1.38)*	1.09 (0.93-1.29)

[a]AOR: adjusted odds ratio.

[b]Ref: reference group.

[c]Income is reported in renminbi ¥. RMB ¥1=US $0.15641.

[d]URBMI: urban-rural resident basic medical insurance.

*P<.05; **P<.01.

Discussion

Principal Findings and Comparison With Prior Work

In this population-based study, we examined the sociodemographic differences in frequency of OHIS and types of online health information sought among the general population in China. This study corroborated the findings of a great deal of previous work in that individuals who were female, better educated, and had a higher household income were positively associated with OHIS [17,18,20,23]. The differences in frequency of OHIS among socioeconomically defined groups implied that inequality in health communication might also exist in China, which was consistent with findings of a previous study from Hong Kong [29]. We also found that living areas had no significant difference between infrequent and frequent online health information seekers, while this difference might exist between seekers and nonseekers [20]. In addition, the sociodemographic characteristics included in the analysis were not consistently associated with the types of online health information sought among the Chinese population.

One unanticipated finding is that respondents who frequently sought online health information were more likely to be middle-aged or older adults. This result deviated from the bulk of previous studies, which showed that younger age groups preferred the internet as a source of health information [17,20,23]. Regardless of sources, however, individuals with older age are more likely to search for health information [4,29]. Moreover, a study from Hong Kong found that older adults were more frequently exposed to health information from instant messaging such as WeChat [42]. Given that our study participants were those who had already searched on the internet for health information, a possible reason for this discrepancy is that once older adults have obtained access to online health information, they are likely to seek it more frequently than are younger individuals, thus suggesting the strong demand for health information among middle-aged or older adults in China. However, additional research effort is needed to verify this hypothesis further. As the population ages in China, the health department of the government and other health organizations should take the internet into account as an effective resource to intervene with the older adults' health behaviors. Given that digital health is not reaching all seniors equally and that this disparity probably results in differences in health outcomes [23], the skills and knowledge of older adults should be reinforced so that more of them can use and benefit from online health information. We also found that the married group had a higher proportion of frequent seekers of online health information, suggesting that married respondents might also seek health information for their family members. A study among German older adults found that internet users were more likely to be married or with a partner [46]. Another study revealed that middle-aged adults were most likely to use the internet to search for information about a condition of a loved one [47]. The association between marital status and the frequency of OHIS may coincide with the association between age and the frequency of OHIS, which may also mean that older people seek out health information for themselves as well as for their partners [47]. Therefore, meeting the health information needs of older adults (eg, information on medical concerns and TCM) is beneficial not only for themselves but also for their partners.

Health science popularization was the most searched for type of online health information among Chinese respondents, and medical concerns was the least searched for type of information. Conversely, earlier literature mostly reported that medical information was the most searched for type of online information [19,30]. This discrepancy may have several causes. On the one hand, the cultural differences between China and other countries, especially Western countries, might have played a role. In recent years, the National Health Commission of China has encouraged the dissemination of health science popularization information in various ways [45]. The increasing popularity of social media and the ever-growing number of official accounts of health science popularization might have attracted many Chinese netizens to follow such information. For instance, by comparing Americans and Hong Kongers, one study revealed that the latter were more likely to trust and use information from social websites [48]. On the other hand, only a minority of respondents in our study had a chronic disease (995/5149, 19.32%) or reported illness in the past 12 months (1074/5149, 20.86%), but they were found more likely to seek information about medical concerns. In addition, the results also showed that respondents over 60 years of age were more likely to seek information about medical concerns, but only 14.29% (736/5149) of the study participants were over 60 years old.

Consistent with previous studies, our study confirmed that people facing health challenges preferred medical information rather than information on healthy behaviors [10,32], which indicated that individuals with more health risks were more eager to improve their current health status. Therefore, people's preferences for health information may be influenced by distinct motivations [49]. More research needs to be conducted to understand Chinese people's motivation in searching for OHIS. Information on healthy behaviors was found to be more likely to be sought by female or better-educated respondents, which was broadly in line with previous studies [50-52]. Our study also found that single respondents were more likely to seek information on healthy behaviors. The sociodemographic differences in preferences for online health information highlight the importance of distinguishing online health information seekers when tailoring and presenting needed information to targeted users.

Information on TCM was more likely to attract attention from older adults, especially older women with chronic diseases. A cross-sectional study focused on older adults in Shanghai indicated that females were more likely to trust and use TCM [53]. Another study showed that having chronic lung disease

was significantly associated with higher TCM practitioner utilization [54]. Indeed, having a chronic condition was identified as one of the main factors that induced people to use TCM in China [55]. As the significant factors found in our study were consistent with those in several previous studies, the preference for online health information about TCM might partly reflect the preference for the use of TCM. Our findings also showed that 37.5% of the respondents had sought TCM information, while a study investigating OHIS behavior of Chinese college students reported that only 15.6% sought TCM information online [44]. Given that TCM is widely accepted and used among the middle-aged and older Chinese population [55], along with the advantages of online health information mentioned before, we believe the provision of TCM-related information on the internet may promote the use of TCM and improve the management of chronic disease in China.

Limitations

Our study has several limitations. First, the content about the motivation for OHIS (eg, for self-care, treating families, and improving lifestyles) and health information sources on the internet (eg, social media and professional websites) were not surveyed, thus resulting in fewer details on the OHIS behavior of Chinese internet users. Second, differences in online health information type classification made our results difficult to compare with those of previous studies. Finally, the research from which this study originated was implemented in 2018. The degree and diversity of internet use has grown since the beginning of the COVID-19 epidemic in 2020. As internet hospitals [56,57], health quick response codes, and antiepidemic science popularization online have been developed and used during the pandemic, the penetration rate of online health information users, especially online medical service users, continues to increase [33]. In addition, the status quo of OHIS in China needs to be further investigated in future research.

Conclusions

This study enriches the current research on the OHIS behavior of the Chinese population. Furthermore, it presents additional knowledge that providers of online health information, government health departments, and other organizations can use. The significant factors found in our study highlight the necessity of identifying the characteristics of the typical online health information users in China, especially age, gender, education level, marital status, and health status, when targeting and developing health interventions by offering health information on the internet. Those who provide health information online should be aware of the needs of specific population groups so that targeted strategies to promote health are presented appropriately.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (#71734003 and #72074087). The funder had no role in the study design; collection, analysis, and interpretation of data; writing of the manuscript; decision to publish; or preparation of the paper.

Authors' Contributions

ZHX, YT, and ZL designed this study; ZL coordinated with the local government and organized the process of data collection; LZ assisted ZL to complete the data entry and quality control; ZHX analyzed the data and drafted the manuscript; YT, ZY, and WCX helped to interpret the results and modified the manuscript, LZ critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. World internet users and 2021 population stats. Internet World Stats. URL: https://www.internetworldstats.com/stats.htm [accessed 2021-04-16]
2. Cherrez Ojeda I, Vanegas E, Torres M, Calderón JC, Calero E, Cherrez A, et al. Ecuadorian cancer patients' preference for information and communication technologies: cross-sectional study. J Med Internet Res 2018 Feb 20;20(2):e50 [FREE Full text] [doi: 10.2196/jmir.8485] [Medline: 29463492]
3. Cash T, Desbrow B, Leveritt M, Ball L. Utilization and preference of nutrition information sources in Australia. Health Expect 2015 Dec;18(6):2288-2295 [FREE Full text] [doi: 10.1111/hex.12198] [Medline: 24798108]
4. Khoong EC, Le GM, Hoskote M, Rivadeneira NA, Hiatt RA, Sarkar U. Health information-seeking behaviors and preferences of a diverse, multilingual urban cohort. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S176-S183 [FREE Full text] [doi: 10.1097/MLR.0000000000001050] [Medline: 31095058]
5. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
6. Peng Y, Yin P, Deng Z, Wang R. Patient-physician interaction and trust in online health community: the role of perceived usefulness of health information and services. Int J Environ Res Public Health 2019 Dec 24;17(1):1-13 [FREE Full text] [doi: 10.3390/ijerph17010139] [Medline: 31878145]

7. 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]
8. Rider T, Malik M, Chevassut T. Haematology patients and the Internet--the use of on-line health information and the impact on the patient-doctor relationship. Patient Educ Couns 2014 Nov;97(2):223-238. [doi: 10.1016/j.pec.2014.06.018] [Medline: 25053474]
9. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
10. Madrigal L, Escoffery C. Electronic health behaviors among us adults with chronic disease: cross-sectional survey. J Med Internet Res 2019 Mar 05;21(3):e11240 [FREE Full text] [doi: 10.2196/11240] [Medline: 30835242]
11. Jin Y, Jing M, Zhang L, Song S, Ma X. Internet access and hypertension management among the elderly population: a nationally representative cross-sectional survey in china. J Med Internet Res 2019 Jan 31;21(1):e11280 [FREE Full text] [doi: 10.2196/11280] [Medline: 30702439]
12. Garfinkle R, Wong-Chong N, Petrucci A, Sylla P, Wexner SD, Bhatnagar S, et al. Assessing the readability, quality and accuracy of online health information for patients with low anterior resection syndrome following surgery for rectal cancer. Colorectal Dis 2019 May;21(5):523-531. [doi: 10.1111/codi.14548] [Medline: 30609222]
13. Storino A, Castillo-Angeles M, Watkins AA, Vargas C, Mancias JD, Bullock A, et al. Assessing the Accuracy and Readability of Online Health Information for Patients With Pancreatic Cancer. JAMA Surg 2016 Sep 01;151(9):831-837. [doi: 10.1001/jamasurg.2016.0730] [Medline: 27144966]
14. Glenton C, Paulsen EJ, Oxman AD. Portals to Wonderland: health portals lead to confusing information about the effects of health care. BMC Med Inform Decis Mak 2005 Mar 15;5:7 [FREE Full text] [doi: 10.1186/1472-6947-5-7] [Medline: 15769291]
15. Schulz PJ, Nakamoto K. Patient behavior and the benefits of artificial intelligence: the perils of "dangerous" literacy and illusory patient empowerment. Patient Educ Couns 2013 Aug;92(2):223-228. [doi: 10.1016/j.pec.2013.05.002] [Medline: 23743214]
16. Varady NH, Dee EC, Katz JN. International assessment on quality and content of internet information on osteoarthritis. Osteoarthritis Cartilage 2018 Aug;26(8):1017-1026. [doi: 10.1016/j.joca.2018.04.017] [Medline: 29758353]
17. Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study. J Med Internet Res 2017 Jun 22;19(6):e216 [FREE Full text] [doi: 10.2196/jmir.6855] [Medline: 28642216]
18. Wang X, Shi J, Lee K. The digital divide and seeking health information on smartphones in asia: a profile of users. J Med Internet Res Preprint posted online September 03, 2020.. [doi: 10.2196/preprints.24086]
19. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2015 Aug 10;22(4):804-814. [doi: 10.1177/1460458215595851] [Medline: 26261218]
20. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a Health Information-Seeking Population and the Current Digital Divide: Cross-Sectional Analysis of the 2015-2016 California Health Interview Survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
21. Asan O, Cooper Ii F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for Health Information Technologies Among US Adults: Analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
22. Massey PM, Langellier BA, Sentell T, Manganello J. Nativity and language preference as drivers of health information seeking: examining differences and trends from a U.S. population-based survey. Ethn Health 2017 Dec;22(6):596-609. [doi: 10.1080/13557858.2016.1244745] [Medline: 27766894]
23. Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, et al. Online Health Information-Seeking Among Older Women With Chronic Illness: Analysis of the Women's Health Initiative. J Med Internet Res 2020 Apr 09;22(4):e15906 [FREE Full text] [doi: 10.2196/15906] [Medline: 32271152]
24. Sherman LD, Patterson MS, Tomar A, Wigfall LT. Use of Digital Health Information for Health Information Seeking Among Men Living With Chronic Disease: Data From the Health Information National Trends Survey. Am J Mens Health 2020;14(1):1557988320901377 [FREE Full text] [doi: 10.1177/1557988320901377] [Medline: 31973642]
25. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2020 Apr 14:1-13. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
26. Chen J, Zhu S. Online Information Searches and Help Seeking for Mental Health Problems in Urban China. Adm Policy Ment Health 2015 May 16;43(4):535-546. [doi: 10.1007/s10488-015-0657-6] [Medline: 25981055]
27. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
28. Tsai M, Chou Y, Lin S, Lin S. Factors associated with adolescents' perspectives on health needs and preference for health information sources in Taiwan. Arch Dis Child 2013 Jan;98(1):9-15. [doi: 10.1136/archdischild-2012-301629] [Medline: 22820106]

29. Wang MP, Viswanath K, Lam TH, Wang X, Chan SS. Social determinants of health information seeking among Chinese adults in Hong Kong. PLoS One 2013;8(8):e73049 [FREE Full text] [doi: 10.1371/journal.pone.0073049] [Medline: 24009729]
30. Ramsey I, Corsini N, Peters MDJ, Eckert M. A rapid review of consumer health information needs and preferences. Patient Educ Couns 2017 Sep;100(9):1634-1642. [doi: 10.1016/j.pec.2017.04.005] [Medline: 28442155]
31. Wong DK, Cheung M. Online Health Information Seeking and eHealth Literacy Among Patients Attending a Primary Care Clinic in Hong Kong: A Cross-Sectional Survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]
32. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
33. The 47th statistical report on China's internet development. China Internet Network Information Center. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202104/P020210420557302172744.pdf [accessed 2021-04-22]
34. Maslen S, Lupton D. "You can explore it more online": a qualitative study on Australian women's use of online health and medical information. BMC Health Serv Res 2018 Dec 03;18(1):916 [FREE Full text] [doi: 10.1186/s12913-018-3749-7] [Medline: 30509261]
35. Freeman JL, Caldwell PHY, Scott KM. The Role of Trust When Adolescents Search for and Appraise Online Health Information. J Pediatr 2020 Jun;221:215-223.e5. [doi: 10.1016/j.jpeds.2020.02.074] [Medline: 32446485]
36. Chae J, Lee C, Jensen JD. Correlates of Cancer Information Overload: Focusing on Individual Ability and Motivation. Health Commun 2016;31(5):626-634. [doi: 10.1080/10410236.2014.986026] [Medline: 26512760]
37. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
38. Tang C, Wu X, Chen X, Pan B, Yang X. Examining income-related inequality in health literacy and health-information seeking among urban population in China. BMC Public Health 2019 Feb 21;19(1):221 [FREE Full text] [doi: 10.1186/s12889-019-6538-2] [Medline: 30791882]
39. Li Z, Zhang L. Poverty and health-related quality of life: a cross-sectional study in rural China. Health Qual Life Outcomes 2020 May 26;18(1):153 [FREE Full text] [doi: 10.1186/s12955-020-01409-w] [Medline: 32456683]
40. Xu W, Li Z, Pan Z, He R, Zhang L. Prevalence and associated factors of self-treatment behaviour among different elder subgroups in rural China: a cross-sectional study. Int J Equity Health 2020 Mar 12;19(1):32 [FREE Full text] [doi: 10.1186/s12939-020-1148-2] [Medline: 32164667]
41. An analysis report of the National Health Services Survey in China, 2013. Center for Health Statistics and Information. 2015. URL: http://www.nhc.gov.cn/ewebeditor/uploadfile/2016/10/20161026163512679.pdf [accessed 2021-04-22]
42. Shen C, Wang MP, Wan A, Viswanath K, Chan SSC, Lam TH. Health information exposure from information and communication technologies and its associations with health behaviors: Population-based survey. Prev Med 2018 Aug;113:140-146. [doi: 10.1016/j.ypmed.2018.05.018] [Medline: 29787775]
43. Woltz DJ, Gardner MK, Kircher JC, Burrow-Sanchez JJ. Relationship between perceived and actual frequency represented by common rating scale labels. Psychol Assess 2012 Dec;24(4):995-1007. [doi: 10.1037/a0028693] [Medline: 22642441]
44. Zhang D, Zhan W, Zheng C, Zhang J, Huang A, Hu S, et al. Online health information-seeking behaviors and skills of Chinese college students. BMC Public Health 2021 Apr 15;21(1):736 [FREE Full text] [doi: 10.1186/s12889-021-10801-0] [Medline: 33858389]
45. Notification of the National Health Commission on printing the guidelines of generating and spreading health science popularization information. National Health Commission of the People's Republic of China. 2015 Aug 11. URL: http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=5fe32b5a1a8243e2bd819f9eeebfd8b1 [accessed 2021-08-15]
46. Quittschalle J, Stein J, Luppa M, Pabst A, Löbner M, Koenig H, et al. Internet Use in Old Age: Results of a German Population-Representative Survey. J Med Internet Res 2020 Nov 23;22(11):e15543 [FREE Full text] [doi: 10.2196/15543] [Medline: 33226351]
47. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. Health Educ Res 2008 Jun;23(3):512-521 [FREE Full text] [doi: 10.1093/her/cyl062] [Medline: 16880222]
48. 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]
49. Dutta-Bergman MJ. Health attitudes, health cognitions, and health behaviors among Internet health information seekers: population-based survey. J Med Internet Res 2004 May 28;6(2):e15 [FREE Full text] [doi: 10.2196/jmir.6.2.e15] [Medline: 15249264]
50. Pollard CM, Pulker CE, Meng X, Kerr DA, Scott JA. Who Uses the Internet as a Source of Nutrition and Dietary Information? An Australian Population Perspective. J Med Internet Res 2015 Aug 26;17(8):e209 [FREE Full text] [doi: 10.2196/jmir.4548] [Medline: 26310192]

51. McCully SN, Don BP, Updegraff JA. Using the Internet to help with diet, weight, and physical activity: results from the Health Information National Trends Survey (HINTS). J Med Internet Res 2013;15(8):e148 [FREE Full text] [doi: 10.2196/jmir.2612] [Medline: 23906945]
52. Visschers VH, Hartmann C, Leins-Hess R, Dohle S, Siegrist M. A consumer segmentation of nutrition information use and its relation to food consumption behaviour. Food Policy 2013 Oct;42:71-80. [doi: 10.1016/j.foodpol.2013.07.003]
53. Xin B, Mu S, Tan T, Yeung A, Gu D, Feng Q. Belief in and use of traditional Chinese medicine in Shanghai older adults: a cross-sectional study. BMC Complement Med Ther 2020 Apr 28;20(1):128 [FREE Full text] [doi: 10.1186/s12906-020-02910-x] [Medline: 32345283]
54. Aw JYH, Yiengprugsawan VS, Gong CH. Utilization of Traditional Chinese Medicine Practitioners in Later Life in Mainland China. Geriatrics (Basel) 2019 Aug 25;4(3):1-15 [FREE Full text] [doi: 10.3390/geriatrics4030049] [Medline: 31450713]
55. Liu T, Li X, Zou Z, Li C. The Prevalence and Determinants of Using Traditional Chinese Medicine Among Middle-aged and Older Chinese Adults: Results From the China Health and Retirement Longitudinal Study. J Am Med Dir Assoc 2015 Nov 01;16(11):1002.e1-1002.e5. [doi: 10.1016/j.jamda.2015.07.011] [Medline: 26344870]
56. Xu X, Cai Y, Wu S, Guo J, Yang L, Lan J, et al. Assessment of Internet Hospitals in China During the COVID-19 Pandemic: National Cross-Sectional Data Analysis Study. J Med Internet Res 2021 Jan 20;23(1):e21825 [FREE Full text] [doi: 10.2196/21825] [Medline: 33417586]
57. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet Hospitals Help Prevent and Control the Epidemic of COVID-19 in China: Multicenter User Profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]

Abbreviations

AOR: adjusted odds ratio
OHIS: online health information seeking
TCM: traditional Chinese medicine

Edited by R Kukafka; submitted 01.06.21; peer-reviewed by Q Yao, X Wang, J Quittschalle; comments to author 04.08.21; revised version received 08.09.21; accepted 29.10.21; published 02.12.21.

Please cite as:
Xiong Z, Zhang L, Li Z, Xu W, Zhang Y, Ye T
Frequency of Online Health Information Seeking and Types of Information Sought Among the General Chinese Population: Cross-sectional Study
J Med Internet Res 2021;23(12):e30855
URL: https://www.jmir.org/2021/12/e30855
doi:10.2196/30855
PMID:34860676

Original Paper

Content and Dynamics of Websites Shared Over Vaccine-Related Tweets in COVID-19 Conversations: Computational Analysis

Iain Cruickshank[1*], BSc, MSc, PhD; Tamar Ginossar[2*], BA, MA, PhD; Jason Sulskis[3*], BSc, MSc; Elena Zheleva[3*], BSc, MSc, PhD; Tanya Berger-Wolf[3,4*], BSc, PhD

[1]Center for Computational Analysis of Social and Organizational Systems, Carnegie Mellon University, Pittsburgh, PA, United States

[2]Department of Communication and Journalism, University of New Mexico, Albuquerque, NM, United States

[3]Department of Computer Science, The University of Illinois at Chicago, Chicago, IL, United States

[4]Translational Data Analytics Institute, The Ohio State University, Colombus, OH, United States

[*]all authors contributed equally

Corresponding Author:
Iain Cruickshank, BSc, MSc, PhD
Center for Computational Analysis of Social and Organizational Systems
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15201
United States
Phone: 1 7192371515
Email: icruicks@andrew.cmu.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2023/1/e43279

Abstract

Background: The onset of the COVID-19 pandemic and the consequent "infodemic" increased concerns about Twitter's role in advancing antivaccination messages, even before a vaccine became available to the public. New computational methods allow for analysis of cross-platform use by tracking links to websites shared over Twitter, which, in turn, can uncover some of the content and dynamics of information sources and agenda-setting processes. Such understanding can advance theory and efforts to reduce misinformation.

Objective: Informed by agenda-setting theory, this study aimed to identify the content and temporal patterns of websites shared in vaccine-related tweets posted to COVID-19 conversations on Twitter between February and June 2020.

Methods: We used triangulation of data analysis methods. Data mining consisted of the screening of around 5 million tweets posted to COVID-19 conversations to identify tweets that related to vaccination and including links to websites shared within these tweets. We further analyzed the content the 20 most-shared external websites using a mixed methods approach.

Results: Of 841,896 vaccination-related tweets identified, 185,994 (22.1%) contained links to specific websites. A wide range of websites were shared, with the 20 most-tweeted websites constituting 14.5% (27,060/185,994) of the shared websites and typically being shared for only 2 to 3 days. Traditional media constituted the majority of these 20 websites, along with other social media and governmental sources. We identified markers of inauthentic propagation for some of these links.

Conclusions: The topic of vaccination was prevalent in tweets about COVID-19 early in the pandemic. Sharing websites was a common communication strategy, and its "bursty" pattern and inauthentic propagation strategies pose challenges for health promotion efforts. Future studies should consider cross-platform use in dissemination of health information and in counteracting misinformation.

(J Med Internet Res 2021;23(12):e29127) doi:10.2196/29127

KEYWORDS

COVID-19; agenda setting; antivaccination; cross-platform; data mining of social media; misinformation; social media; Twitter; vaccinations; vaccine hesitancy

Introduction

Misinformation over social media contributes to the global growth in vaccine hesitancy. The World Health Organization (WHO) defined vaccine hesitancy as the "delay in acceptance or refusal of vaccines despite availability of vaccine services" [1] and declared it as one of the top 10 global health challenges in 2019, just before the outbreak of COVID-19 [2]. The "infodemic" that ensued as a response to the pandemic enhanced concerns about the rise of vaccine hesitancy in the 21st century [3] and about the key role that social media play in dissemination of vaccine-related misinformation [3]. Vaccine hesitancy discourse on social media does not represent just individual behavior. These messages are often part of concentrated disinformation efforts, intentionally promoted by nationalist right-wing politicians [4] and specific antivaccination leaders and "celebrities" [5,6]. Researchers also documented the efforts of foreign governments aiming to destabilize democratic processes by eroding the public trust in social institutions, including public health sources and the mainstream media [7-9]. With the COVID-19 pandemic, an increase was documented in the volume of misinformation on social media, including antivaccination propaganda [3,10-14]. Early in the pandemic, public health officials were concerned that the pandemic and the response to it led to worries about a global decrease in access to and acceptance of childhood and other vaccinations. In addition, public health efforts centered on developing a vaccine as a central strategy for ending the pandemic. Therefore, public trust in the safety and efficacy of vaccinations was considered paramount. Examining discourse on vaccination on social media is key to understanding public sentiments and to identifying the specific strategies used by different users, including antivaccination advocates, to erode trust in vaccinations.

Twitter has documented importance in setting the public and political players' agendas [15], including in vaccinations [16,17]. According to agenda-setting theory [18], issues that are presented in the media frequently and prominently gain perceived salience by audiences. This salience is important as it impacts political agenda and policy making. According to Langer and Gruber (page 314) [19], "Unless an issue gets into the political agenda, it will not be discussed, debated in the legislature or acted upon by the government...news coverage is an important factor in making policy change more likely." The theory was created to describe processes of *traditional media*, and specifically of legacy news. Traditional media, or *old media*, refer to the centralized mass media institutions that predated the information age, including print, television, radio broadcasting, studio-produced movies, and large advertising firms, among others [20,21]. In contrast to traditional media communication that was based on one-way technologies, new media are based on interactive and largely decentralized computer technologies [22], with the internet as the delineating telecommunication network [23].

Contrary to predictions about the death of traditional media in the age of the internet, and particularly following the rise of social media, studies indicate that they remain important. Traditional media have documented synergy with social media that amplify their mutual impact on the agenda in intermediate agenda-setting processes [19]. Research on the role of mass media during pandemics is fragmented, but a recent study that used computational methods in exploring the role of the media in covering pandemics revealed limited coverage of governmental health sources and frameworks [24]. This media coverage is important in influencing community behaviors. For instance, a study of COVID-19 coverage and behavior in Italy showed that the frames used by the news media influenced changes in community mobility significantly more than the effect of the number of daily death reports [25]. Moreover, whereas the most common source type for COVID-19 information seeking online was media outlets followed by governmental sources, governmental sources were the most likely to meet medical benchmark criteria for quality [26].

Studies on Twitter's role in political agenda setting revealed an intermediate effect, in which the agendas of traditional media and Twitter were dissimilar, but exerted mutual influences [15]. However, past studies did not examine such processes in the context of vaccination-related tweets. Due to Twitter's role in spreading health information and misinformation [27], understanding vaccination-related content and agenda-setting processes on this social platform can advance public health research and knowledge and inform future interventions.

Studies that used surveys to examine individual beliefs and intentions yielded important information on how audiences make sense of novel vaccines in the face of emerging pandemics, including their use of mental frameworks from previously known vaccines [28]. However, surveys are limited due to human recall and by access to participants. It is, therefore, pertinent to analyze the content and dynamics of social media that individuals create, share, and consume. New computational approaches to analyzing big data allow for analysis of communication about vaccination over social media in unprecedented ways [29-33].

Several studies documented the role of Twitter in disseminating vaccine-related misinformation prior to the current pandemic [8,34-38]. However, few studies examined patterns of this discourse. Notably, Twitter discourse about vaccination was reported as featuring heterogeneous conversations that were not dominated by particular subjects, sources, or users. Information sources that were tweeted frequently included health-specific sites, national media, medical organizations, and digital news aggregators [16]. A more recent COVID-19–related study [13] revealed that the largest single topic of Twitter conversations included comparisons between COVID-19 and influenza. Propagation of misinformation was observed in both previously known and new vaccine-opposing sources [13]. The study established that known sources of vaccine communication

continued to engage in the topic early in the pandemic. However, it was limited to analysis of tweets that were tweeted over the course of one day. To get a more holistic picture of these conversations, it is important to examine communication about vaccination as part of Twitter's COVID-19 discourse over extended time frames.

An additional lacuna in research on vaccine-related communication on social media involves its focus on single social media platforms. Research documented that most social media users use multiple media sources and social media platforms [39], and often go back and forth between different platforms [40]. It is, therefore, important to explore cross-platform use. Analysis of links to websites and the website domains that are shared on tweets can provide information about such cross-platform use and spread of information sources. Specifically, including a URL in a tweet allows readers to link to the website. For instance, most tweets that responded to misinformation with cross-platform links during Hurricanes Harvey and Irma focused on debunking misinformation and used news source URLs in their response [41]. Hence, such cross-platform use can serve to share information during a time of crisis. Examination of website sharing in COVID-19 conversations documented both the importance of traditional news sources and the propensity for virality of low-quality sources. Whereas low-quality information sources were tweeted at higher rates compared to high-quality health sources, traditional news sources were shared at a much higher rate than other sources [42]. Moreover, shared websites within COVID-19 Twitter conversations revealed users' political stance [43], thus lending more support to the close links between health and political debates during the pandemic.

These previous studies underscored the potential importance of cross-platform information sharing on Twitter. Analyzing both the URLs and the domains shared as external content on Twitter can provide insights into the type of specific content and information sources included in social media messages about vaccination. Despite this importance, vaccine-related cross-platform use over Twitter received limited scholarly attention. Examination of links to websites shared within vaccination-related tweets early in the COVID-19 pandemic can enrich knowledge by gaining a broader understanding of this communication. Empirical implications of this knowledge include informing strategies for evidence-based vaccine-related message dissemination over social media. Moreover, it can shed light on the role of traditional media in the era of social media, as well as on how social media are used in cahoots with vaccine-related communication and their dynamics over time. Finally, in view of the documented role of inauthentic propagation of vaccine-related content on Twitter [8,37] and in COVID-19–related discussions [44], studies should go beyond typologies of vaccine-related content [45] and understand strategies employed in the spread of this content. New social cybersecurity methods [46] can aid in such examinations.

The goal in this study was to examine website sharing in vaccine-related tweets posted to COVID-19 conversations in the 20 weeks following the declaration of the pandemic. Specifically, we analyzed tweets that were posted from February 1 (two days after the WHO declared the outbreak of COVID-19 to be a Public Health Emergency of International Concern) through June 23, 2020, and sought to examine the magnitude, temporal patterns, and content of websites shared within these tweets. This study will provide unique contributions to theory and practice. Our examined time frame took place prior to the development of the vaccine and the implementation of COVID-19 vaccination campaigns. Therefore, tweets posted during that period can indicate the degree to which vaccinations were included in COVID-19–related discourse from its inception. It will also reveal the information sources that were promoted through cross-platform link sharing. These findings have the potential to indicate the effectiveness of official health sources in leading the agenda as health information providers and the prominence of vaccine-opposing sources. It can also uncover some of the tactics of the vaccination-opposing movement over time and in response to this new, unexpected, global health threat. This understanding is important for advancing theories about the role of social media in public health crises, as well as for informing future policies, interventions, and dissemination of health information to address audiences' informational and emotional needs.

Given the importance of vaccination-related discourse and website sharing within COVID-19 Twitter conversations, including an understanding of the sources of vaccination-related information and their spread, we posed the following research questions.

First, as this is, to our knowledge, the first study to examine external content sharing in the context of vaccination in early COVID-19 conversations, we were interested in understanding the magnitude of external content, the degree to which vaccinations were featured in conversations about COVID-19 early in the pandemic, and the prevalence and dynamics of website sharing within these tweets.

Therefore, we posed the first research question: What are the prevalence and dynamics of vaccination-related tweets, including website sharing, posted between February 1 and June 23, 2020, as part of COVID-19 conversations, as evident in the number of these tweets over time?

We were further interested in learning about agenda-setting processes that are demonstrated in this relatively new social media strategy of information source promotion. As websites' domains, such as television networks, vlogs, or individual social media accounts, represent specific information sources, we aimed to learn about these sources and their characteristics. Specifically, in view of the importance of information sources in public health communication, we sought to identify the sources of the websites that were shared most prominently in the early months of the pandemic in tweets about vaccination that were posted to COVID-19 Twitter conversations.

Therefore, the second research question was posed: What are the characteristics of the 20 most-shared website domains?

In addition to the information sources, we were interested in exploring the content of the most-shared information, as evident in the 20 most-tweeted websites in our data set. The prominence of these websites can stem from users being activated by the content and their desire to share it. However, specific spread

strategies and coordinated efforts might also drive this prominence. Therefore, we wanted to examine both the content of the websites and the specific information that was shared in this cross-platform modality, as well as their propagation.

Therefore, the third research question was posed: What characterizes the content and spread of the 20 most-shared websites?

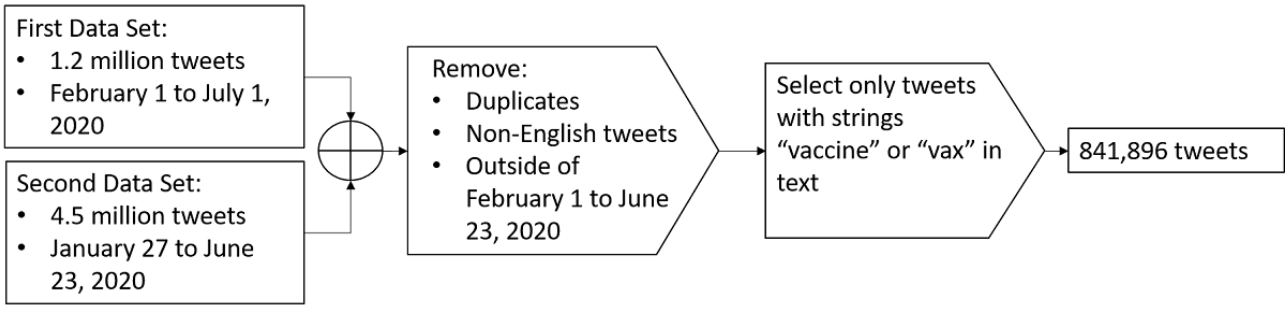
Methods

Data

The analysis encompassed two different data sets of COVID-19–related tweets. The first data set was based on a collection of tweet IDs gathered using general COVID-19–related keywords, such as "coronavirus" and "Wuhancoronavirus" [47]. We used "hydration" [48], a process of gathering all the pertinent information about each tweet into the JSON format file [49] via the Twitter search application programming interface (API) [50], on all of these tweet IDs. This process only populated data from tweets that were available on Twitter at the time of hydration and would exclude banned users or deleted tweets. This data set had around 1 million tweets. The second data set included approximately 4.5 million COVID-19–related tweets collected from January 29 to June 23, 2020, using Twitter's streaming API [51,52]. Since these tweets were collected in a streaming fashion, as they were tweeted, it allowed for analysis of some tweets that were otherwise no longer available on Twitter.

Both data sets were then filtered to include only the dates of overlap (February 1 to June 23, 2020) and to remove any duplicated tweets across the data sets. We then filtered the data to include only English-language tweets. Given our interest in tweets about vaccination, each data set was filtered using the substrings "vax" and "vaccin." This process ensured that the tweets included in our analysis referred to vaccinations. The resulting data set contained 841,896 English-language tweets. Since our focus was on analyzing content available to users rather than the identity of users, we did not attempt to distinguish between human users and machine accounts (ie, bots) [53]. Figure 1 displays a graphical flowchart for the data selection and exclusion process.

Figure 1. Data set combination, filtering, and exclusion process.



URL Extraction

We first extracted the website URLs from the JSON "entities" object of each tweet in order to get the original URL rather than the version automatically shortened by Twitter. URLs that were still shortened were unshortened to their original form using an API [54]. To preprocess the URLs for analysis, we then removed all URL query terms from all domains.

Content Analysis

In addition to computational methods, mixed methods content analysis was conducted in analyzing the 20 most-tweeted websites to identify the source of the websites and their content. First, the sources of the 20 most-tweeted URLs and the dates they were posted were recorded. The content of the 20 most-tweeted URLs was coded using inductive qualitative methods using the constant comparative method [55-57]. This design, which prioritized quantitative methods, is consistent with an *explanatory sequential* design. This design involves implementing quantitative research methods first, followed by qualitative methods, with the aim of explaining the quantitative methods [58]. Mixed methods research is particularly appropriate in studying complex social phenomena, and this approach was appropriate, as the quantitative approach was not deemed sufficient [58-60]. The qualitative research approach is suitable for exploratory studies when researchers are unable to use theory to produce hypotheses or theoretical-driven prediction [61].

The qualitative analysis followed a multistep iterative process. A coauthor with expertise in mixed methods research created initial codes and recorded memos. The initial coding involved line-by-line detailed reading of the data, aimed at understanding the different views and actions described in the different URLs and approaching coding in an inductive manner while remaining open to different potential theoretical directions emerging from the data [62]. During the second phase of the analysis, focused coding was conducted. The focused coding entailed coding of the significant and frequent themes that emerged during the initial coding. Focused coding was helpful in synthesizing and conceptualizing the data and the research [62], while also remaining cognizant of the different sources generating the content. Comparisons of statements and incidents were noted within and across the different URLs. At the second stage, previous research and categorizations were considered in addition to the texts at hand, and these informed the categories of the URLs' framing. The first category included content that overtly advanced doubts about at least one of the following: vaccines' efficacy, vaccines' safety, and the motives of those who fund, develop, and/or test them (ie, vaccine-opposed). Conversely, the second category captured content that featured the efficacy and/or safety of COVID-19 vaccines. The third

category included content that focused on advances in development of vaccines, including news on the development of specific vaccines and related scientific breakthroughs. In this third theme, coders also annotated whether the advancements that were reported were based on meaningful developments or whether they reflected anecdotal information and unfounded claims. The fourth category related to content that highlighted political aspects of vaccination, including portraying political processes as influencing vaccine development and availability to the public. This political content was further coded to capture whether vaccination was, in fact, the focus of the overall content. In addition, the coders noted whether content in any category could have increased distrust in vaccination by using implicit cues that casted doubts on the integrity of the process, those developing vaccines, or decisions and decision makers. For instance, a news story that announced that a COVID-19 vaccine was developed in 3 hours was coded as including "vaccine-opposing" sentiment, as it was judged to be increasing concerns about a vaccine that was developed so rapidly and, therefore, likely to reduce trust in its safety and efficacy.

In the third stage of the qualitative analysis, the overall theme of "politicizing vaccination" emerged. A graduate student with training in qualitative research followed this process independently by coding each of the 20 URLs using inductive coding first and then coding by the previous categories. They then provided a quotation from each URL to support the coding. As a final check on consistency of results, we evaluated the intercoder agreement for all of the URLs between the initial coding and the third stage of qualitative analysis. The use of constant comparative analysis [62] helped us develop our analytical categories, including attention to contradictions.

Twitter Spread Analysis

To examine the particular Twitter spread strategies that were used to propagate the most-tweeted websites, we have applied social cybersecurity methods to identify coordinated link sharing and flooding (or spamming) of the websites by tweeters [46,63]. Specifically, we analyzed all the tweets that shared the top 20 most-tweeted websites. First, we excluded retweets and removed the following text from the remaining original tweets: mentions, URLs, trailing white space, and formatting characters (ie, "\n"). Then, we recorded the number of tweets that included a URL to each website, unique users that tweeted the website, unique texts across all of the tweets that contained the website, tweets that featured a website that were tweeted within an hour of the first tweet of that same website, and nonreply mentions, as well as the range of days between the first and last tweet that included the website.

Results

Overview

Our aims were to understand the magnitude, dynamics over time, content, sources, and spread of websites shared within vaccine-related tweets as part of COVID-19 Twitter conversations. The first research question centered on the prevalence and dynamics of vaccination-related tweets, as well as website sharing posted as part of COVID-19 conversations over time. The analysis revealed that these conversations demonstrated an overall growth in tweets that related to vaccination. It also showed that website sharing had distinct patterns compared to overall tweets about vaccination. As mentioned above, our data set contained a total of 841,896 tweets. As seen in Figure 2, tweets about vaccination spiked in March and again in June. A corresponding spike in tweets that were retweeted was observed in March. In contrast, a more modest increase was observed in March in website sharing, followed by a leveling in April. In addition, sharing websites and the diversity in unique websites shared increased over time, indicating that Twitter users spread more external content from a greater number of sources as the scope and scale of the pandemic increased in the early months of the pandemic.

Our analysis revealed that 1 in 5 of the 841,896 tweets (n=185,994, 22.1%) contained at least one website. A total of 1 in 4 of the 524,998 users (n=128,408, 24.5%) tweeted at least one website. In comparison, only 19.4% (n=163,743) of all tweets contained at least one hashtag, and 23.0% (n=120,699) of users tweeted at least one hashtag. Additionally, 85.2% (n=717,150) of all tweets contained a mention of another user's account (ie, using the "@" symbol to refer to another Twitter account), and 87.4% (n=459,038) of users tweeted at least one mention in a tweet. Of the mentions, only 12.1% (n=87,097) were replies (ie, when a user was directly replying back to the tweet of another user). Figure 3 displays the rates of usage for these different social media artifacts.

A total of 1 out of 5 of the tweets that mentioned vaccination (185,994/841,896, 22.1%) included links to websites. These websites included 11,311 unique website domains. Most domains (n=6962, 61.6%) were tweeted with only one unique website tweeted.

Figure 2. Counts of unique tweets, users, and website domains for all of the vaccination tweets.

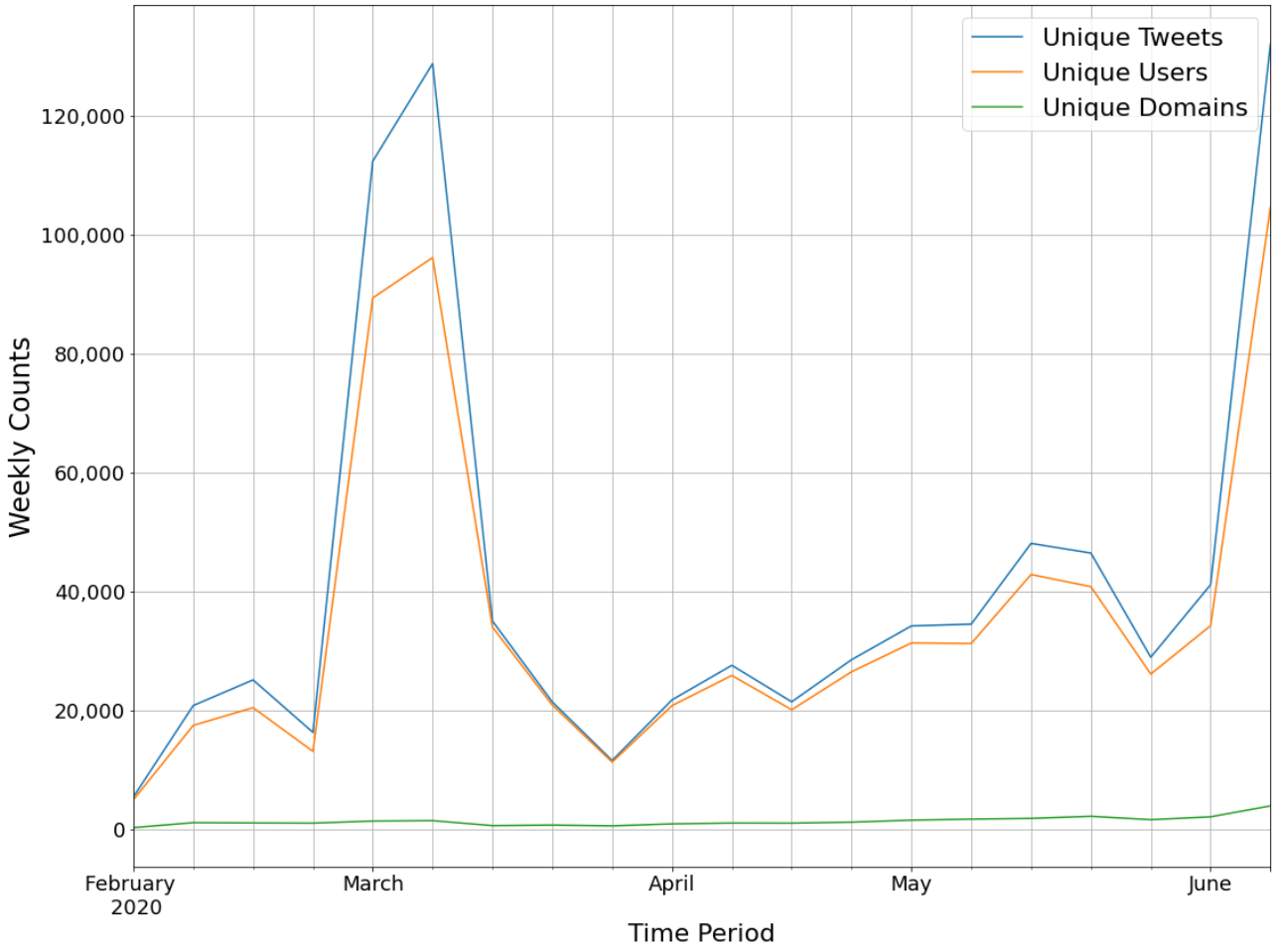
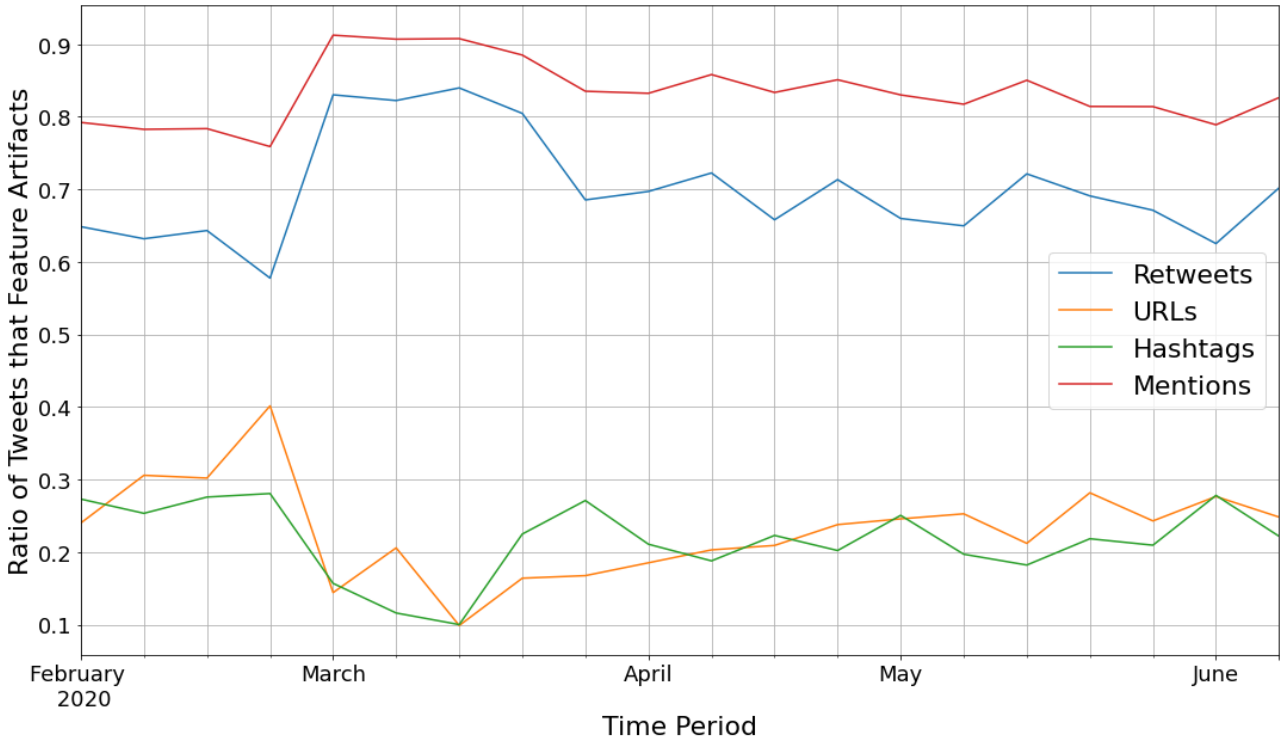


Figure 3. Rates of retweets, hashtags, mentions, and website sharing in all of the vaccination tweets.



Analysis of Website Domains

Research question 2 explored the 20 most-shared website domains, with a focus on the role of legacy media, social media, and public health sources. As displayed in Table 1 [64-83], the majority (n=14, 70%) of the 20 most-tweeted website domains consisted of traditional news media, including news organizations, newspapers, and television networks. However, the most-tweeted website domain was Raw Story, a US online tabloid that is classified as progressive. Another tabloid, the New York Post, was ranked as the 19th most-tweeted website domain. The only social media platform included was YouTube, which was the third most-tweeted domain. Similarly, the only official governmental and/or health source was the US Centers for Disease Control and Prevention (CDC), which was ranked as the 20th most-tweeted website domain.

Websites associated with these 20 domains constituted 8.25% (n=6244) of all of the 75,642 websites tweeted. Two of the most-tweeted website domains—Raw Story and The Jerusalem Post—had most of their tweets from just one news story each. Notably, the top two most–individually tweeted websites were in this category, indicating that "viral" tweets can increase a domain's popularity.

On average, each domain had 4.26 (SD 24.4) websites per domain, with a mode of 1 website, indicating that many of the domains only ever had one website associated with them. In total, 62.7% (n=6963) of all domains in the data had only one unique website associated with that domain. An exception was YouTube, which had the greatest number of unique websites of any domain, followed by the CDC website.

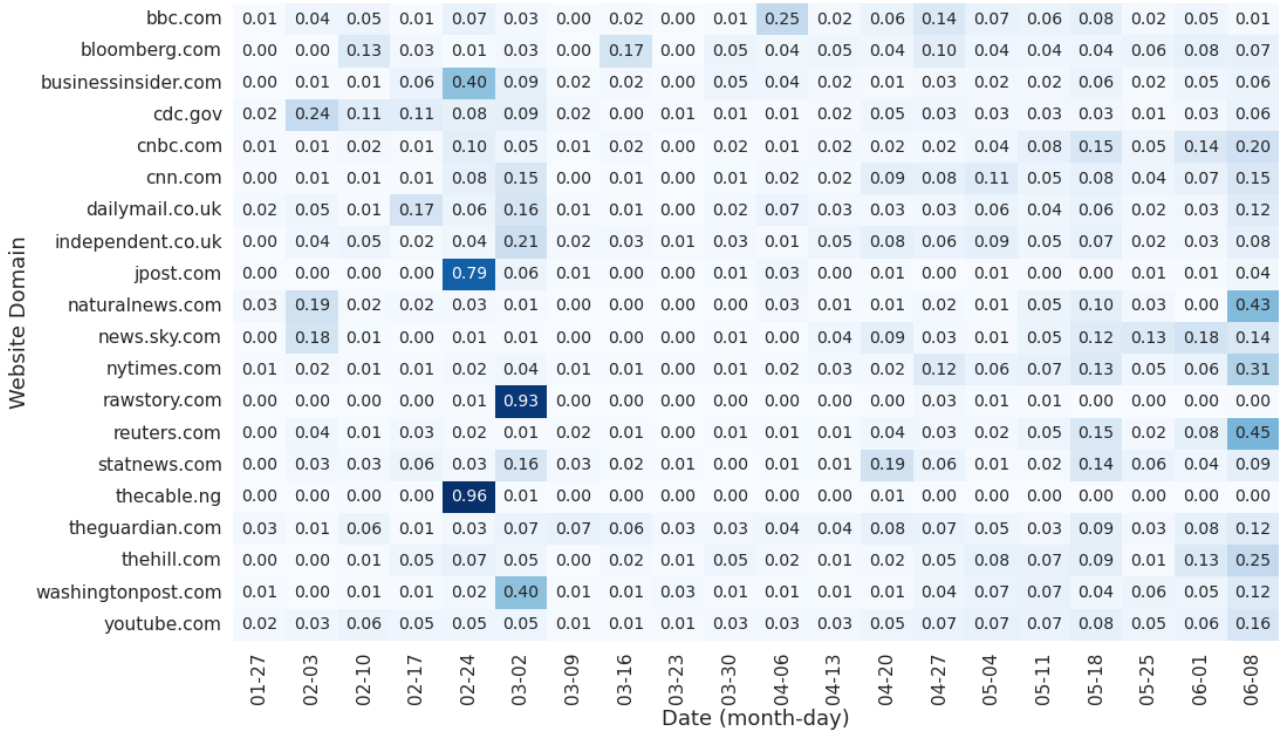
Table 1. Top 20 most-tweeted website domains.

Website domain	Tweets, n	Percentage of all tweets with website URLs that originate from website domain, %	Unique websites per domain[a], n	Type of domain and country of origin
Raw Story [64]	13,261	7.1	101	US online tabloid
Reuters [65]	4347	2.3	577	International news organization
YouTube [66]	4106	2.2	2013	International, US-based social media platform
The Guardian [67]	3167	1.7	364	UK newspaper
The Jerusalem Post [68]	3140	1.7	92	Israeli newspaper
Bloomberg [69]	2374	1.3	173	International, US-based news agency
CNBC [70]	2161	1.2	282	US television channel
The Daily Mail [71]	2102	1.1	291	UK newspaper
CNN [72]	1888	1.0	302	Multinational, US-based television channel
The New York Times [73]	1716	0.9	332	US newspaper
The Cable [74]	1642	0.9	19	Nigerian digital newspaper
STAT [75]	1552	0.8	122	Health-oriented US news website
Business Insider [76]	1447	0.8	203	US financial news website
The Washington Post [77]	1348	0.7	209	US newspaper
BBC [78]	1320	0.7	106	UK public service broadcast organization
Sky News [79]	1301	0.7	147	UK television news channel
The Independent [80]	1280	0.7	172	UK newspaper
The Hill [81]	1270	0.7	152	US newspaper
New York Post [82]	1148	0.6	152	US conservative-leaning tabloid
Centers for Disease Control and Prevention [83]	1139	0.6	435	US government health organization

[a]This is the number of unique websites that originate from the higher-level domain. For example, a news website can have several unique websites representing different news stories that all come from the same single news website domain.

The temporal analysis of the patterns of the 20 most-tweeted domains documented three distinct peaks of domain usage, as presented in Figure 4. The peaks included The Jerusalem Post and The Cable from February 20 to 24 and Raw Story on March 4. These spikes in domain usage were the result of tweeting a particular website from that domain. Specific stories that became "viral" underlined domain activity. Tweeted domains were typically shared over one week, or even one day. This pattern can, therefore, be characterized as "bursty," as opposed to having different websites associated with specific domains tweeted consistently across longer periods of time.

Figure 4. The 20 most-used domains overall by the number of tweets featuring each domain. Counts are normalized within each time period (week).



In contrast to the majority of the domains that were characterized by "bursty" activity, some domains, such as YouTube and The Guardian, had consistent usage over time. The 20 most–persistently tweeted domains are summarized in Table 2 [65-67,69-73,75-78,80,82-88].

As Table 2 shows, YouTube was the most–persistently tweeted domain. Out of the 20 domains, the 12 (60%) that were tweeted most persistently over time were news organizations, and 2 (10%) included official US federal health organizations. Both the CDC and the National Institutes of Health (NIH) were included in this list, indicating the persistent sharing of content of these domains. Similarly, in addition to YouTube, Instagram was a second social media platform that made the list, as well as Google.

Table 2. The 20 most–persistently tweeted website domains.

Home domain	Type of domain	Days that a website from the domain was tweeted at least once, %
YouTube [66]	Social media platform	77
The Guardian [67]	UK newspaper	76
The Daily Mail [71]	UK newspaper	72
Centers for Disease Control and Prevention [83]	US federal health organization	71
CNBC [70]	US television news channel	71
The New York Times [73]	US newspaper	70
Reuters [65]	International news agency	70
Bloomberg [69]	US business newspaper	67
STAT [75]	Health-oriented US news website	66
Business Insider [76]	US financial news website	65
Instagram [84]	Social media platform	65
The Independent [80]	UK newspaper	65
CNN [72]	Multinational, US-based television channel	64
Google [85]	Multinational technology company	64
New York Post [82]	US tabloid	63
BBC [78]	UK public service broadcast organization	61
MSN [86]	Web portal by Microsoft	61
The Washington Post [77]	US newspaper	61
BBC [87]	UK public service broadcast organization	60
NCBI (National Center for Biotechnology Information), NIH (National Institutes of Health) [88]	US federal health organization	59

Analysis of Websites

The third research question explored the content and dynamics of spread of the 20 most-tweeted websites. Table 3 [89-108] displays the 20 most-tweeted individual websites within the data.

The 20 most-tweeted websites comprised around 13% of all the tweets containing websites (n=185,994) in the data set. The most-tweeted website, which was by Raw Story, accounted for almost half of these, with 6.6% (n=12,201) of the overall number of tweets containing websites. As Table 3 indicates, the majority (n=11, 55%) of the 20 most-tweeted websites were tweeted by traditional media sources or news organizations, and the sources of 6 other websites (30%) were tabloid and digital-only newspapers and websites. One fake news website, a petition website for right-wing causes, and another domain were the sources of three more websites. The latter two were the only nonnews sources among the top 20 most-tweeted websites.

As an overarching theme, the qualitative analysis indicated that the content of these websites demonstrated politicization of vaccination. The single most-tweeted website was by the digital tabloid Raw Story, which described Republicans blocking the COVID-19 bill to avoid posing limits on pharmaceutical companies' charges for the vaccine. This politicization of vaccination was evident in other websites. As seen in Table 3, 9 out of 20 (45%) websites (#1, #4-6, #8, #9, #15, #17, and #20) focused on political aspects of the vaccine and its development. In 2 of them (#6 and #15), vaccines were mentioned merely as a minor issue. Moreover, 2 of these featured cues that could increase, or were related to, distrust in vaccination. These included a Fox Business story (#14) claiming that a vaccine was developed in 3 hours in collaboration with China and was funded by the Gates Foundation, and a BBC News story (#18) that reported on comments by a top health official in France that called to test the COVID-19 vaccine in Africa. Although the WHO was cited as denying these comments, the story, overall, featured distrust in vaccine development, provided direct citations of celebrities that responded to the possibility of testing in Africa, and referred to controversial testing of HIV medications on prostitutes in Africa. Similarly, the STAT story (#20) that covered Dr Rick Bright's departure from directing the Biomedical Advanced Research and Development Authority, without stating a reason for his departure, could have led audiences to have reduced trust given their portrayals of plots that left much to the imagination. Other frames included criticism of then-President Trump's competence in combating the pandemic following comments or decisions that he made, including suggesting using a flu vaccine to prevent COVID-19 (#8 and #15) and pulling out of a national effort to speed vaccine development (#17).

Table 3. The 20 most-tweeted websites.

Rank	Title of webpage	Source; type	Topic	Coding	Tweets, n	Date in 2020
1	GOP blocking coronavirus bill — because it limits how much drugmakers can charge for a vaccine: Report [89]	Raw Story; US progressive-leaning tabloid	Republicans block coronavirus bill because it limits how much pharmaceutical companies can charge	4. Political focus: the main focus is on the political processes and motives related to vaccination	12,201	March 3
2	Israeli scientists: 'In a few weeks, we will have coronavirus vaccine' [90]	The Jerusalem Post; Israeli newspaper	Israeli scientists are close to developing a COVID-19 vaccine	3b. News on vaccine development that proved unfounded	2780	April 13
3	Israeli researchers announce breakthrough on coronavirus vaccine [91]	The Cable; Nigerian online newspaper	Israeli scientists are close to developing a COVID-19 vaccine	3b. News on vaccine development that proved unfounded	1507	February 9
4	EU sets out plans for advance orders of coronavirus vaccines [92]	The Irish Times; Irish newspaper	Britain will not be included in European COVID-19 vaccine supplies	4. Political focus: the main focus is on the political processes and motives related to vaccination	865	June 11
5	UK will not participate in EU's coronavirus fast track vaccine scheme [93]	The New European; UK pro-Europe newspaper	Britain will not be included in European COVID-19 vaccine supplies	4. Political focus: The main focus is on the political processes and motives related to vaccination	789	June 12
6	James Clapper refuses to testify to Congress in person 'until there's a COVID vaccine' [94]	True Pundit; US fake news website	Former Director of National Intelligence refused to testify until there is a vaccine	4. Political focus: the main focus is on the political processes and motives related to vaccination	530	May 15
7	AstraZeneca agrees to supply Europe with 400 million doses of COVID-19 vaccine [95]	Reuters; international news organization	AstraZeneca signed a contract to supply 400 million doses of COVID-19 vaccine to Europe	3a. News on vaccine development that proved founded	521	June 13
8	Read Pedagogy of the Oppressed by Paulo Freire (tweet by Joshua Potash, liberal leaning with 146,000 followers) [96]	Twitter; US-based social media platform	"Trump thinks we should use the flu vaccine to defend against coronavirus. We could not be in worse hands."	4. Political focus: the main focus is on the political processes and motives related to vaccination	512	March 2
9	Trump's baffling coronavirus vaccine event [97]	Washington Post; US newspaper	Commentary on Trump's interactions with vaccine makers	4. Political focus: the main focus is on the political processes and motives related to vaccination	454	March 3
10	PETITION: No to mandatory vaccination for the coronavirus [98]	Life Petitions; petitions website	Petition to prevent mandatory COVID-19 vaccination	1. Content that overtly advances doubts regarding vaccines' efficacy and safety and the motives of those who fund, develop, and/or test them	434	January 23
11	Meet the all-female team working to create a COVID-19 vaccine in Maryland [99]	WJLA: ABC News; local DC news affiliate	Team working on developing COVID-19 vaccine	3b. News on vaccine development that proved unfounded	428	February 28
12	US scientists have completed a coronavirus vaccine, Texas-based genetic engineering company claims [100]	The Daily Mail; UK newspaper	Texas-based scientists reported completion of COVID-19 vaccine development	3b. News on vaccine development that proved unfounded	391	February 20
13	Israeli scientists: 'In a few weeks, we will have coronavirus vaccine' [101]	The Jerusalem Post; Israeli newspaper (mobile version)	Development of COVID-19 vaccine in Israel	3b. News on vaccine development that proved unfounded	387	April 13

Rank	Title of webpage	Source; type	Topic	Coding	Tweets, n	Date in 2020
14	California lab says it discovered coronavirus vaccine in 3 hours [102]	Fox Business; US television channel	Development of COVID-19 in California in 3 hours, funded by Gates, with China	3b. News on vaccine development that proved unfounded	381	February 13
15	Experts baffled as Trump asks why they can't just use flu vaccines to prevent coronavirus [103]	Indy100; UK online newspaper	Experts are baffled at Trump suggestion to use flu vaccine	4. Political focus: the main focus is on the political processes and motives related to vaccination	377	March 3
16	COVID-19 vaccine shipped, and drug trials start [104]	TIME; US news magazine	Moderna vaccine shipped, and trials start	3a. News on vaccine development that proved founded	363	February 25
17	Trump removes US from global initiative to develop coronavirus treatments and vaccines [105]	Raw Story; US tabloid	Trump removes United States from global initiative to develop COVID-19 treatment	4. Political focus: the main focus is on the political processes and motives related to vaccination	348	February 25
18	Coronavirus: Africa will not be testing ground for vaccine, says WHO[a] [106]	BBC News; UK broadcast organization	WHO says Africa will not be a testing ground for COVID-19 vaccine	3b. News on vaccine development that proved unfounded	325	April 6
19	Here's why Obamacare would likely make any coronavirus vaccine free for patients — and prove critical in fighting the disease [107]	Business Insider; US online media company	COVID-19 vaccine will be available for free thanks to the Affordable Care Act	2. Provaccination: the content focuses on the efficacy and/or safety of COVID-19 vaccines, as a way to end the pandemic	319	February 29
20	Director of US agency key to vaccine development leaves role suddenly amid coronavirus pandemic [108]	STAT; US health news website	Rick Bright steps down	4. Political focus: the main focus is on the political processes and motives related to vaccination	300	April 21

[a]WHO: World Health Organization.

Of the 9 (45%) websites that covered news about advances in vaccine development, only 2 of the developments that were covered proved founded (#7 and #16 covered AstraZeneca and Moderna, respectively). TIME reported on the rollout of Moderna vaccine clinical trials (#16), stood out as the only website that provided medical framing of the content, and included explanations regarding the vaccine's mechanism of using messenger RNA (mRNA). In contrast, 7 (35%) stories covered advancements that, at the time of analysis, were unfounded or did not come to fruition (#2, #3, #11-14, and #17). Notably, 3 of these included news on development of an Israeli vaccine that allegedly was 3 days away from the finish line and 90 days from approval. This information came from the Israeli Science and Technology Minister, and the sources were an Israeli newspaper in English and a Nigerian digital newspaper. Of the coverage of vaccine development that did not reach the market, only the report about the Novavax vaccine (#11) included information about the phase of the trial. The report on Greffex (#12) included some scientific information on the technology of the vaccine and a timeline that emphasized the lengthy process.

Only 1 (5%) website expressed explicit opposition to vaccination (#10). It consisted of a petition to block mandatory vaccination, which was included in a website domain that promoted petitions for conservative, right-wing causes. Similarly, only 1 (5%) website (#19), by Business Insider, a US online media company, provided provaccination framing by positioning vaccines as the way to end the pandemic.

Table 4 [89-108] presents information on the original tweets that included each of the 20 most-tweeted websites.

Table 4. Tweets spreading the 20 most-tweeted websites.

Title of webpage	Website	Unique tweets with URL, n	Unique texts of tweets, n	Unique users that tweeted URL, n	Mentions in tweets with URL, mean (SD)	Retweets of tweets with URL, n	Days between first and last tweet of URL, n	Tweets within first hour of first tweet, n
GOP blocking coronavirus bill — because it limits how much drugmakers can charge for a vaccine: Report [89]	Raw Story	838	437	786	0.31 (1.61)	11,363	12	129
Israeli scientists: 'In a few weeks, we will have coronavirus vaccine' [90]	The Jerusalem Post	710	352	612	0.33 (1.14)	2070	78	20
Israeli researchers announce breakthrough on coronavirus vaccine [91]	The Cable	31	20	26	0.71 (1.54)	1476	3	4
EU sets out plans for advance orders of coronavirus vaccines [92]	The Irish Times	78	41	77	0.21 (0.43)	787	2	1
UK will not participate in EU's coronavirus fast track vaccine scheme [93]	The New European	50	34	48	0.08 (0.3)	739	2	2
James Clapper refuses to testify to Congress in person 'until there's a COVID vaccine' [94]	True Pundit	20	13	20	0 (0)	510	21	1
AstraZeneca agrees to supply Europe with 400 million doses of COVID-19 vaccine [95]	Reuters	62	38	57	0.21 (0.57)	459	1	5
Read Pedagogy of the Oppressed by Paulo Freire [96]	Twitter	112	111	110	0.16 (0.49)	400	2	7
Trump's baffling coronavirus vaccine event [97]	The Washington Post	351	197	339	0.19 (0.75)	103	66	19
PETITION: No to mandatory vaccination for the coronavirus [98]	Life Petitions	241	90	156	0.50 (2.32)	193	37	3
Meet the all-female team working to create a COVID-19 vaccine in Maryland [99]	WJLA: ABC News	56	33	56	0.16 (0.49)	372	18	2
US scientists have completed a coronavirus vaccine, Texas-based genetic engineering company claims [100]	The Daily Mail	135	64	133	0.67 (0.75)	256	26	22
Israeli scientists: 'In a few weeks, we will have coronavirus vaccine' [101]	The Jerusalem Post (mobile version)	110	55	104	0.79 (3.99)	277	51	1
California lab says it discovered coronavirus vaccine in 3 hours [102]	Fox Business	225	123	215	0.24 (0.69)	156	19	8
Experts baffled as Trump asks why they can't just use flu vaccines to prevent coronavirus [103]	Indy100	5	2	4	0.2 (0.4)	372	0	3
COVID-19 vaccine shipped, and drug trials start [104]	TIME	111	81	109	0.43 (1.16)	252	63	1

Title of webpage	Website	Unique tweets with URL, n	Unique texts of tweets, n	Unique users that tweeted URL, n	Mentions in tweets with URL, mean (SD)	Retweets of tweets with URL, n	Days between first and last tweet of URL, n	Tweets within first hour of first tweet, n
Trump removes US from global initiative to develop coronavirus treatments and vaccines [105]	Raw Story	17	13	17	0.52 (1.24)	331	1	3
Coronavirus: Africa will not be testing ground for vaccine, says WHO [106]	BBC	18	14	17	0.38 (1.16)	307	1	2
Here's why Obamacare would likely make any coronavirus vaccine free for patients — and prove critical in fighting the disease [107]	Business Insider	230	127	222	0.64 (1.05)	89	6	4
Director of US agency key to vaccine development leaves role suddenly amid coronavirus pandemic [108]	STAT	24	11	24	0.58 (0.57)	276	6	6

The Twitter propagation statistics of some of the websites show attempts at spreading the websites through inauthentic means. Specifically, a high percentage of the tweets that included links to The Cable website (#3) and to the STAT website (#20) had a high percentage of retweets only, an indicator of spreading by flooding or spamming of a website on a social media platform in an effort to get the website artificially trending and, consequently, to give more exposure to other social media users. In addition, the text of over half of the tweets that included the 20th most-tweeted website, STAT, was identical, which is indicative of coordinated, inauthentic link sharing. Evidence of such coordinated inauthentic link sharing was also present in tweets that included links to the Life Petitions website (#10), which also had relatively few unique tweet texts compared to the number of tweets of the website and the number of unique users that tweeted that website.

Discussion

Principal Findings

This study is the first to examine the prevalence, dynamics, and content of websites shared in vaccination-related tweets. We focused on tweets that were part of COVID-19 conversations over 20 weeks, following the WHO's announcement of COVID-19 as a pandemic until June 23, 2020. The main finding of this study is the use of a cross-platform strategy for promoting politicization of COVID-19 vaccination well before the rollout of these vaccines. This politicization of content, promotion of unfounded "advancements" in vaccine development, and coverage of unsettling political plots that left much unexplained are likely, in turn, to contribute to a decrease both in the public's knowledge of the science behind vaccine development and effectiveness and its trust in vaccination. Future studies should investigate the impact of exposure to this coverage.

Whereas previous research on the topic typically focused on the degree to which Twitter discussions reflected specific vaccine sentiments, our study indicates the politicization of the topic, which was shared by both progressive-leaning sources and content (eg, the Raw Story online tabloid) and legacy media (eg, The Washington Post) as well as by right wing–leaning sources, including legacy media (eg, The Jerusalem Post) and fake news sources (eg, True Pundit).

The websites that were tweeted represented diverse communication sources, with traditional news media making the top shared domains. Both the prominence of legacy news media in websites shared and the emergence of nontraditional media outlets, such as tabloids, vlogs, and other social media, exemplify processes of intermediate agenda setting in the new media environment. These processes were previously documented in political content [15], and this study extends them to this health context. Along with "rehashing" legacy news content, as was evident in the majority of the sources shared, Twitter has also given rise to nontraditional, digital-only content. These nontraditional sources typically reached salience in terms of website sharing when a story they published became viral. For instance, Raw Story, a digital tabloid [109], featured the most-tweeted website in its story of Republicans blocking a bill in order to protect pharmaceutical companies from limitations on vaccine-related profits. The salience of nontraditional sources demonstrates an intermedia agenda-setting process that provides a platform for individuals who were previously blocked from entering the elite spaces to disseminate their messages [15,110]. Twitter "has become an important platform for eloquent and media-savvy people outside the traditional political, economic, or academic elites" [15]. Our study extends this line of research to intermedia agenda setting in vaccine-related conversations. The content of the URLs shared over Twitter represented, to a great degree, an alternative agenda. In this agenda, stories that advanced political motives that went beyond the issue of vaccination were featured prominently. They represented both opposition to Trump's US presidential administration at the time of data collection and right-wing populist views, including vaccine-opposed content. Similarly, while some of the news coverage about vaccine development stood the test of time, a

few stories that reported on vaccine advancement were inaccurate, such as the Israeli development of a vaccine. This type of coverage is likely to increase public doubt regarding news in general and scientific news in particular. Agenda setting by non-legacy media sources has both theoretical and practical implications for public health efforts. Unfortunately, in this context, these sources are also used by Twitter users associated with misinformation, conspiracy theories, and vaccine opposing messages.

In this new media environment, official health sources like the CDC and the NIH have had some success in disseminating their information, indicated by their inclusion in the lists of the most-tweeted domains (ie, CDC) and the most–consistently shared domains (ie, CDC and NIH). This is important, as governmental sources have been shown to provide credible, high-quality information compared to other sources, including the media [42]. However, producing the most-tweeted unique websites, in other words, "becoming viral," proved more challenging, at least in the context of cross-platform sharing. The importance of trustworthy sources that provide scientifically sound information to the public is heightened at a time of a pandemic. Our findings revealed that the CDC had a salient role in the vaccine conversation on Twitter as the 20th most-tweeted website domain in our sample. The CDC was also second only to YouTube in providing a large number of different websites, which indicated that it provided diverse information that was deemed important by Twitter users who felt the need to share this information over Twitter. In comparison, no NIH-specific unique link to a website was included in the top 20 most-tweeted websites, but its domain was one of the 20 most–consistently tweeted domains. Although it is also possible that the CDC, NIH, and other public health sources exuded additional influence via links to traditional media shared over Twitter, such influence was not evident in the 20 most-tweeted websites.

Given the prominence of traditional media with their established gatekeeping, checks, and balances, it is not surprising that most stories shared in websites did not include obvious vaccine-opposing content. This finding is consistent with previous content analyses of tweets, reporting that vaccine-opposing content comprised a minority of the overall discussions on Twitter [16,111,112]. While this is encouraging from a public health perspective, it is important to remember that the impact of misinformation might still be significant, particularly in view of the social network nature of Twitter that often broadcasts to specific groups [36] and the need for herd immunity in maximizing the effects of vaccinations [112].

Our findings also point at the salience of international content. In addition to large, global media institutions like Reuters or CNN, and US-based newspapers and tabloids, some British newspapers, most notably The Guardian, were heavily tweeted in our data. Moreover, both the Israeli newspaper, The Jerusalem Post, and the Nigerian digital newspaper, The Cable, were included in list of the most-shared websites thanks to the viral story about an alleged Israeli COVID-19 vaccine. Similar to other smaller media organizations that were heavily tweeted, these two foreign, small newspapers demonstrated the opportunity of small players to advance their agenda in this new media environment by becoming "viral" through provision of sensational narratives. In the case of the Nigerian newspaper, inauthentic targeted attempts to spread this story contributed to its popularity, demonstrating the importance of deliberate manipulation in the new social media environment. It is likely that this rapidly changing environment, characterized by a "bursty" pattern of website sharing and the need to continuously provide new and sensational narratives as well as information and inauthentic spread strategies, poses unique challenges for governmental and official health sources. In addition, social media cross-platform sharing was evident in the prominence of YouTube as the third most-shared domain, as well as the most–consistently shared website over time. Future studies should further explore the content included in the different platforms, as well as users' interpretation of this content and their motivation to engage in sharing it.

Finally, these findings are important in revealing patterns of propagation of this external content and these links. Despite the known presence of bots and other inauthentic propagation strategies of vaccination-related content on Twitter [8,37], and regarding COVID-19 [44], previous studies typically focused on analyzing the content of tweets in attempts to identify misinformation [45]. Such studies are important in advancing the knowledge and theories concerning the content to which users are exposed. However, the propagation strategies of this content should also be understood and considered. For instance, interventions to block such content should consider the propagation strategies. Given silos in current studies, social cybersecurity methods, to our knowledge, were not previously applied to vaccination-related discourse. Our study, therefore, is important in providing an opportunity to explore propagation of vaccine-related content over Twitter by spread of external content.

These results have important implications that can inform interventions, policies, and future research. At the most basic level, our findings indicate that sharing links to websites is a common strategy in Twitter conversations on the topic. In fact, shared websites were more common than hashtags, which have become synonymous with Twitter. Hashtags are frequently researched due to their use in creating discussion communities on social media [113,114]. It is, therefore, significant that in the context of the vaccine discussions we examined, external websites were featured as frequently as hashtags.

Our analysis also revealed that websites were tweeted in a "bursty" pattern, indicating heterogeneity of a large number of sources, stories, and topics shared. The results regarding the increased number and diversity of external links shared at the time of data collection are consistent with other studies that documented the ebbs and flows of the "infodemic." A recent study suggested that this increase was motivated by both uncertainty and state-sponsored propaganda [115]. COVID-19 is used as vector to propagate misinformation and disinformation by foreign governments. In addition, it provides a highly uncertain information environment in which fact-checking is difficult. The authors emphasized the importance of constant, reliable medical information provided by governmental sources. While we concur with this suggestion, it is important to note that our findings also point at the challenges of such public

health response. Given the numbers, diversity of sources, and dynamics of the topic that are constantly evolving, such response would require significant efforts and resources [115].

Strengths and Limitations

The strengths of this study stem from its analysis of a large data set that was collected at a historically important period. Moreover, we employed a triangulation of computational methods and human coding to study a previously unexplored communication strategy within vaccine discourse on Twitter. However, this study is not without limitations. First, our tweets were collected by searching common vaccine-related keywords and hashtags. While these keywords and hashtags were identified following an extensive literature review and analysis of tweets by multiple research teams, it is possible that some emerging keywords and hashtags were not included. Future studies could apply additional computational methods, such as the Analysis of Topic Model Networks [116].

Additional limitations are grounded in our focus on tweets in English and on a specific time frame. Future studies should expand research to include additional languages and time frames, particularly during and following the rollout of the COVID-19 vaccine. In addition, some vaccine-related tweets, particularly those advancing vaccine-opposing messages, were deleted by Twitter by the time of analysis. Hence, the actual number of antivaccination tweets shared might be higher than what we were able to report, and their content might be somewhat different from what was collected. Moreover, we have focused on vaccine-related tweets that were part of the COVID-19 Twitter conversations. Although our data set is unique in including all related tweets rather than a sample, our findings do not apply to vaccine-related discourse on Twitter that was not part of the pandemic discourse. Moreover, our study focused on the first 20 weeks of the pandemic. Future studies should compare our findings regarding website sharing with similar content following the implementation of the COVID-19 vaccination campaigns globally. In addition, we focused on the content and propagation of the 20 most-tweeted websites and domains, and these findings might not apply to other links shared in this data set.

Conclusions

These findings are important in advancing understanding of website sharing in vaccine-related tweets, its use, and its dynamics. The analysis revealed that Twitter users share websites as part of their vaccine messages in COVID-19 conversations and that some of this sharing revealed inauthentic, deliberate attempts to spread this content. Our data included tweets that were posted in the first 5 months of the pandemic and showed that vaccine-related tweets were prominent in the pandemic-related Twitter discourse from its inception. Future research should examine the following months, as it is likely that with the advances in vaccine development, these conversations have increased in frequency and perhaps included different information sources.

The findings of this study pave the way for future studies that would answer additional questions. First and foremost, future research should expand the scope of this study by examining websites shared after June 2020, especially as new COVID-19 vaccines were approved and disseminated, and as information became available about their safety and efficacy. Given that our findings encompass the period prior to the approval and dissemination of the specific COVID-19 vaccines, they shed light on early communication on the topic rather than the specific risks and benefits of these vaccines.

In view of the global importance of the pandemic and vaccinations, future studies should also expand the scope of analysis to include additional languages other than English. In addition, it is important to consider, measure, and analyze additional aspects and implications of our work. For instance, to date, studies did not explore the impact of visual content on vaccine-related messages over social media. Future studies should expand the scope of this study's analysis by exploring the visual content of vaccine-related tweets, websites, and of YouTube videos on the topic and the visual impact on propagation of the content over social networks. Similarly, future studies should explore the content and propagation of additional URLs in addition to the 20 most-tweeted websites explored in this study.

Finally, we call for hypothesis-driven communication interventions that would not only measure how and why antivaccination messages propagate over social media [117] and attempt to correct misinformation [118], but would also attempt to prevent this propagation and advance scientifically accurate content instead. Such future interventions should not focus on one social media platform, but should instead consider and integrate cross-platform use for message sharing.

Acknowledgments

This research was, in part, funded by the Omar N Bradley Foundation through a General Omar N Bradley Research Fellowship in Mathematics Grant (IC) and was conducted with support from the University of New Mexico Women in STEM Faculty Development Fund (TG). We would like to thank Melanie Catron, Zohar Griffith, Kenneth Hafer, Paulina Majewska, and Rachel Taylor for their help in coding the data.

Conflicts of Interest

None declared.

References

1. Dubé E, Gagnon D, Nickels E, Jeram S, Schuster M. Mapping vaccine hesitancy--Country-specific characteristics of a global phenomenon. Vaccine 2014 Nov 20;32(49):6649-6654 [FREE Full text] [doi: 10.1016/j.vaccine.2014.09.039] [Medline: 25280436]
2. Dubé E, MacDonald NE. How can a global pandemic affect vaccine hesitancy? Expert Rev Vaccines 2020 Oct;19(10):899-901. [doi: 10.1080/14760584.2020.1825944] [Medline: 32945213]
3. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
4. Żuk P, Żuk P. Right-wing populism in Poland and anti-vaccine myths on YouTube: Political and cultural threats to public health. Glob Public Health 2020 Jun;15(6):790-804. [doi: 10.1080/17441692.2020.1718733] [Medline: 31964228]
5. Bradshaw AS, Treise D, Shelton SS, Cretul M, Raisa A, Bajalia A, et al. Propagandizing anti-vaccination: Analysis of Vaccines Revealed documentary series. Vaccine 2020 Feb 18;38(8):2058-2069. [doi: 10.1016/j.vaccine.2019.12.027] [Medline: 31980194]
6. Larson HJ, Cooper LZ, Eskola J, Katz SL, Ratzan S. Addressing the vaccine confidence gap. Lancet 2011 Aug 06;378(9790):526-535. [doi: 10.1016/S0140-6736(11)60678-8] [Medline: 21664679]
7. Ernst N, Blassnig S, Engesser S, Büchel F, Esser F. Populists prefer social media over talk shows: An analysis of populist messages and stylistic elements across six countries. Soc Media Soc 2019 Jan 16;5(1):1-14 [FREE Full text] [doi: 10.1177/2056305118823358]
8. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
9. Walter D, Ophir Y, Jamieson KH. Russian Twitter accounts and the partisan polarization of vaccine discourse, 2015–2017. Am J Public Health 2020 May;110(5):718-724. [doi: 10.2105/ajph.2019.305564]
10. Shahi GK, Dirkson A, Majchrzak TA. An exploratory study of COVID-19 misinformation on Twitter. Online Soc Netw Media 2021 Mar;22:100104 [FREE Full text] [doi: 10.1016/j.osnem.2020.100104] [Medline: 33623836]
11. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
12. 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]
13. Jamison AM, Broniatowski DA, Dredze M, Sangraula A, Smith MC, Quinn SC. Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic' on Twitter. Harv Kennedy Sch Misinformation Rev 2020 Sep;1:1-24 [FREE Full text] [doi: 10.37016/mr-2020-38] [Medline: 34368805]
14. Romer D, Jamieson KH. Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]
15. Rogstad I. Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. J Inf Technol Polit 2016 May 27;13(2):142-158. [doi: 10.1080/19331681.2016.1160263]
16. 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]
17. McNeill A, Harris PR, Briggs P. Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. Front Public Health 2016;4:26 [FREE Full text] [doi: 10.3389/fpubh.2016.00026] [Medline: 26942174]
18. Coleman R, McCombs M, Shaw D, Weaver D. Agenda setting. In: Wahl-Jorgensen K, Hanitzsch T, editors. The Handbook of Journalism Studies. New York, NY: Routledge; 2009:147-160.
19. Langer AI, Gruber JB. Political agenda setting in the hybrid media system: Why legacy media still matter a great deal. Int J Press Polit 2020 Jun 10;26(2):313-340. [doi: 10.1177/1940161220925023]
20. Becker B, Wehner J. Electronic networks and civil society: Reflections on structural changes in the public sphere. In: Ess C, editor. Culture, Technology, Communication: Towards an Intercultural Global Village. Albany, NY: State University of New York Press; 2001:67-85.
21. Gross L. Out of the mainstream: Sexual minorities and the mass media. J Homosex 1991 May 13;21(1-2):19-46. [doi: 10.1300/j082v21n01_04]
22. Harrison TM, Barthel B. Wielding new media in Web 2.0: Exploring the history of engagement with the collaborative construction of media products. New Media Soc 2009 Feb 01;11(1-2):155-178. [doi: 10.1177/1461444808099580]
23. Greenstein S. How the Internet Became Commercial. Princeton, NJ: Princeton University Press; Oct 20, 2015.
24. Ophir Y. Coverage of epidemics in American newspapers through the lens of the crisis and emergency risk communication framework. Health Secur 2018;16(3):147-157. [doi: 10.1089/hs.2017.0106] [Medline: 29870279]
25. Ophir Y, Walter D, Arnon D, Lokmanoglu A, Tizzoni M, Carota J, et al. The framing of COVID-19 in Italian media and its relationship with community mobility: A mixed-method approach. J Health Commun 2021 Mar 04;26(3):161-173. [doi: 10.1080/10810730.2021.1899344] [Medline: 33787462]

26. Sajjadi NB, Shepard S, Ottwell R, Murray K, Chronister J, Hartwell M, et al. Examining the public's most frequently asked questions regarding COVID-19 vaccines using search engine analytics in the United States: Observational study. JMIR Infodemiology 2021;1(1):e28740 [FREE Full text] [doi: 10.2196/28740] [Medline: 34458683]
27. 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 Jan;107(1):e1-e8. [doi: 10.2105/ajph.2016.303512]
28. Ophir Y, Jamieson KH. Intentions to use a novel Zika vaccine: The effects of misbeliefs about the MMR vaccine and perceptions about Zika. J Public Health (Oxf) 2018 Dec 01;40(4):e531-e537 [FREE Full text] [doi: 10.1093/pubmed/fdy042] [Medline: 29554290]
29. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
30. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: A quantitative analysis of YouTube videos. Hum Vaccin Immunother 2018 Jul 03;14(7):1654-1659 [FREE Full text] [doi: 10.1080/21645515.2018.1454572] [Medline: 29553872]
31. Wardle C, Singerman E. Too little, too late: Social media companies' failure to tackle vaccine misinformation poses a real threat. BMJ 2021 Jan 21;372:n26. [doi: 10.1136/bmj.n26] [Medline: 33478950]
32. Karafillakis E, Martin S, Simas C, Olsson K, Takacs J, Dada S, et al. Methods for social media monitoring related to vaccination: Systematic scoping review. JMIR Public Health Surveill 2021 Feb 08;7(2):e17149 [FREE Full text] [doi: 10.2196/17149] [Medline: 33555267]
33. Himelboim I, Xiao X, Lee DKL, Wang MY, Borah P. A social networks approach to understanding vaccine conversations on Twitter: Network clusters, sentiment, and certainty in HPV social networks. Health Commun 2020 May;35(5):607-615. [doi: 10.1080/10410236.2019.1573446] [Medline: 31199698]
34. Tavoschi L, Quattrone F, D'Andrea E, Ducange P, Vabanesi M, Marcelloni F, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from September 2016 to August 2017 in Italy. Hum Vaccin Immunother 2020 May 03;16(5):1062-1069 [FREE Full text] [doi: 10.1080/21645515.2020.1714311] [Medline: 32118519]
35. Featherstone JD, Ruiz JB, Barnett GA, Millam BJ. Exploring childhood vaccination themes and public opinions on Twitter: A semantic network analysis. Telematics Inform 2020 Nov;54:101474. [doi: 10.1016/j.tele.2020.101474]
36. Cossard A, De Francisci Morales G, Kalimeri K, Mejova Y, Paolotti D, Starnini M. Falling into the echo chamber: The Italian vaccination debate on Twitter. In: Proceedings of the 14th International AAAI Conference on Web and Social Media. 2019 Presented at: 14th International AAAI Conference on Web and Social Media; June 8-11, 2019; Virtual p. 130-140 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7285/7139
37. Yuan X, Schuchard RJ, Crooks AT. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. Soc Media Soc 2019 Sep 04;5(3):1-12 [FREE Full text] [doi: 10.1177/2056305119865465]
38. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Soc Sci Med 2017 Oct;191:168-175 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.041] [Medline: 28926775]
39. Kulshrestha J, Zafar M, Noboa L, Gummadi K, Ghosh S. Characterizing information diets of social media users. In: Proceedings of the 9th International AAAI Conference on Web and Social Media. 2015 Presented at: 9th International AAAI Conference on Web and Social Media; May 26-29, 2015; Oxford, UK URL: https://arxiv.org/pdf/1704.01442.pdf
40. Tandoc EC, Lou C, Min V. Platform-swinging in a poly-social-media context: How and why users navigate multiple social media platforms. J Comput Mediat Commun 2019;24(1):21-35. [doi: 10.1093/jcmc/zmy022]
41. Hunt K, Wang B, Zhuang J. Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: A case study on shelters and ID checks. Nat Hazards 2020 May 27;103(1):861-883. [doi: 10.1007/s11069-020-04016-6]
42. Singh L, Bode L, Budak C, Kawintiranon K, Padden C, Vraga E. Understanding high- and low-quality URL sharing on COVID-19 Twitter streams. J Comput Soc Sci 2020 Nov 27:1-24 [FREE Full text] [doi: 10.1007/s42001-020-00093-6] [Medline: 33263092]
43. Cruickshank IJ, Carley KM. Characterizing communities of hashtag usage on Twitter during the 2020 COVID-19 pandemic by multi-view clustering. Appl Netw Sci 2020;5(1):66 [FREE Full text] [doi: 10.1007/s41109-020-00317-8] [Medline: 32953977]
44. Ferrara E. #COVID-19 on Twitter: Bots, conspiracies, and social media activism. ArXiv. Preprint posted online on April 20, 2020 [FREE Full text]
45. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and extending a typology to identify vaccine misinformation on Twitter. Am J Public Health 2020 Oct;110(S3):S331-S339. [doi: 10.2105/ajph.2020.305940]
46. Carley KM. Social cybersecurity: An emerging science. Comput Math Organ Theory 2020 Nov 16:1-17 [FREE Full text] [doi: 10.1007/s10588-020-09322-9] [Medline: 33223952]

47. 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]
48. Hydrator. GitHub. URL: https://github.com/DocNow/hydrator [accessed 2021-11-17]
49. Pezoa F, Reutter JL, Suarez F, Ugarte M, Vrgoč D. Foundations of JSON schema. In: Proceedings of the 25th International Conference on World Wide Web. 2016 Apr 11 Presented at: 25th International Conference on World Wide Web; April 11-15, 2016; Montreal, QC p. 263-273. [doi: 10.1145/2872427.2883029]
50. Search Tweets: Standard v1.1. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets [accessed 2021-11-17]
51. Huang B. Learning User Latent Attributes on Social Media [doctoral thesis]. Pittsburgh, PA: Carnegie Mellon University; 2020 May. URL: https://kilthub.cmu.edu/articles/thesis/Learning_User_Latent_Attributes_on_Social_Media/12307157 [accessed 2021-11-24]
52. Consuming streaming data. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data [accessed 2021-11-17]
53. Alothali E, Zaki N, Mohamed EA, Alashwal H. Detecting social bots on Twitter: A literature review. In: Proceedings of the 13th International Conference on Innovations in Information Technology. 2019 Jan 10 Presented at: 13th International Conference on Innovations in Information Technology; November 18-19, 2018; Al Ain, United Arab Emirates p. 175-180. [doi: 10.1109/innovations.2018.8605995]
54. Quickstart. Unshortenit. 2018. URL: https://unshortenit.readthedocs.io/en/latest/quickstart.html [accessed 2021-11-17]
55. Glaser BG. The constant comparative method of qualitative analysis. Soc Probl 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
56. Fram S. The constant comparative analysis method outside of grounded theory. Qual Rep 2015 Jan 14(18):1-25. [doi: 10.46743/2160-3715/2013.1569]
57. Hewitt-Taylor J. Use of constant comparative analysis in qualitative research. Nurs Stand 2001;15(42):39-42. [doi: 10.7748/ns2001.07.15.42.39.c3052] [Medline: 12212430]
58. Creswell JW. Advanced mixed methods research designs. In: Tashakkori A, Teddlie C, editors. Handbook of Mixed Methods in Social & Behavioral Research. Thousand Oaks, CA: SAGE Publications; Dec 01, 2004:240.
59. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications; 2006.
60. Creswell JW, Fetters MD, Ivankova NV. Designing a mixed methods study in primary care. Ann Fam Med 2004;2(1):7-12 [FREE Full text] [doi: 10.1370/afm.104] [Medline: 15053277]
61. Pluye P, Hong QN. Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45. [doi: 10.1146/annurev-publhealth-032013-182440] [Medline: 24188053]
62. Charmaz K. Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis. Thousand Oaks, CA: SAGE Publications; 2006.
63. Giglietto F, Righetti N, Rossi L, Marino G. Coordinated link sharing behavior as a signal to surface sources of problematic information on Facebook. In: Proceedings of the 11th International Conference on Social Media and Society. 2020 Presented at: 11th International Conference on Social Media and Society; July 22-24, 2020; Virtual p. 85-91. [doi: 10.1145/3400806.3400817]
64. Raw Story. URL: https://www.rawstory.com/ [accessed 2021-11-18]
65. Reuters. URL: https://www.reuters.com/ [accessed 2021-11-18]
66. YouTube. URL: https://www.youtube.com/ [accessed 2021-11-18]
67. The Guardian. URL: https://www.theguardian.com/international [accessed 2021-11-08]
68. The Jerusalem Post. URL: https://www.jpost.com/ [accessed 2021-11-18]
69. Bloomberg. URL: https://www.bloomberg.com/ [accessed 2021-11-18]
70. CNBC. URL: https://www.cnbc.com/ [accessed 2021-11-18]
71. The Daily Mail. URL: https://www.dailymail.co.uk/ [accessed 2021-11-18]
72. CNN. URL: https://www.cnn.com/ [accessed 2021-11-18]
73. The New York Times. URL: https://www.nytimes.com/ [accessed 2021-11-18]
74. The Cable. URL: https://www.thecable.ng/ [accessed 2021-11-18]
75. STAT. URL: https://www.statnews.com/ [accessed 2021-11-18]
76. Business Insider. URL: https://www.businessinsider.com/ [accessed 2021-11-18]
77. The Washington Post. URL: https://www.washingtonpost.com/ [accessed 2021-11-18]
78. BBC. URL: https://www.bbc.com/ [accessed 2021-11-18]
79. Sky News. URL: https://news.sky.com/ [accessed 2021-11-18]
80. The Independent. URL: https://www.independent.co.uk/ [accessed 2021-11-18]
81. The Hill. URL: https://thehill.com/ [accessed 2021-11-18]
82. The New York Post. URL: https://nypost.com/ [accessed 2021-11-18]
83. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/ [accessed 2021-11-18]

84. Instagram. URL: https://www.instagram.com/ [accessed 2021-11-18]
85. Google. URL: https://www.google.com/ [accessed 2021-11-18]
86. MSN. URL: https://a.msn.com/en-ca/ [accessed 2021-11-18]
87. BBC. URL: https://www.bbc.co.uk/ [accessed 2021-11-18]
88. NCBI, NIH. URL: https://www.ncbi.nlm.nih.gov/ [accessed 2021-11-18]
89. Chapman M. GOP blocking coronavirus bill — because it limits how much drugmakers can charge for a vaccine: Report. Raw Story. 2020 Mar 03. URL: https://www.rawstory.com/2020/03/gop-blocking-coronavirus-bill-because-it-limits-how-much-drugmakers-can-charge-for-a-vaccine-report/ [accessed 2021-11-18]
90. Jaffe-Hoffman M. Israeli scientists: 'In a few weeks, we will have coronavirus vaccine'. The Jerusalem Post. 2021 Mar 21. URL: https://www.jpost.com/HEALTH-SCIENCE/Israeli-scientists-In-three-weeks-we-will-have-coronavirus-vaccine-619101 [accessed 2021-11-18]
91. Owolabi F. Israeli researchers announce breakthrough on coronavirus vaccine. The Cable. 2020 Feb 29. URL: https://www.thecable.ng/israeli-researchers-announce-breakthrough-in-developing-coronavirus-vaccine [accessed 2021-11-18]
92. O'Leary N. EU sets out plans for advance orders of coronavirus vaccines. The Irish Times. 2020 Jun 11. URL: https://www.irishtimes.com/news/world/europe/eu-sets-out-plans-for-advance-orders-of-coronavirus-vaccines-1.4276885 [accessed 2021-11-18]
93. UK will not participate in EU's coronavirus fast track vaccine scheme. The New European. 2020 Jun 12. URL: https://web.archive.org/web/20200616054015/https://www.theneweuropean.co.uk/top-stories/uk-government-left-out-of-coronavirus-scheme-due-to-brexit-1-6697267 [accessed 2021-11-24]
94. Johnston A. James Clapper refuses to testify to Congress in person 'until there's a COVID vaccine'. True Pundit. 2020 May 15. URL: https://truepundit.com/james-clapper-refuses-to-testify-to-congress-in-person-until-theres-a-covid-vaccine/ [accessed 2021-11-18]
95. Reuters staff. AstraZeneca agrees to supply Europe with 400 million doses of COVID-19 vaccine. Reuters. 2020 Jun 13. URL: https://www.reuters.com/article/us-health-coronavirus-vaccines-idUSKBN23K0HW [accessed 2021-11-18]
96. Read Pedagogy of the Oppressed by Paulo Freire. Twitter. 2020 Mar 02. URL: https://twitter.com/JoshuaPotash/status/1234661401488052226 [accessed 2021-11-18]
97. Blake A. Trump's baffling coronavirus vaccine event. The Washington Post. 2020 Mar 03. URL: https://www.washingtonpost.com/politics/2020/03/03/trumps-baffling-coronavirus-vaccine-event/ [accessed 2021-11-18]
98. PETITION: No to mandatory vaccination for the coronavirus. Life Petitions. 2021 May 18. URL: https://lifepetitions.com/petition/no-mandatory-vaccine-for-covid-19 [accessed 2021-11-18]
99. Sanchez V. Meet the all-female team working to create a COVID-19 vaccine in Maryland. WJLA: ABC News. 2020 Feb 28. URL: https://wjla.com/news/health/all-female-team-covid-19-coronavirus-vaccine-maryland [accessed 2021-11-18]
100. Rahhal N. US scientists have completed a coronavirus vaccine, Texas-based genetic engineering company claims. The Daily Mail. 2020 Feb 20. URL: https://www.dailymail.co.uk/health/article-8026293/US-scientists-completed-coronavirus-vaccine.html [accessed 2021-11-18]
101. Jaffe-Hoffman M. Israeli scientists: 'In a few weeks, we will have coronavirus vaccine'. The Jerusalem Post. 2020 Feb 27. URL: https://m.jpost.com/HEALTH-SCIENCE/Israeli-scientists-In-three-weeks-we-will-have-coronavirus-vaccine-619101 [accessed 2021-11-18]
102. Fordham E. California lab says it discovered coronavirus vaccine in 3 hours. Fox Business. 2020 Feb 13. URL: https://www.foxbusiness.com/technology/california-lab-coronavirus-vaccine-3-hours [accessed 2021-11-18]
103. Bergman S. Experts baffled as Trump asks why they can't just use flu vaccines to prevent coronavirus. Indy100. 2020 Mar 03. URL: https://www.indy100.com/news/coronavirus-trump-vaccine-flu-us-symptoms-9372566 [accessed 2021-11-18]
104. Park A. COVID-19 vaccine shipped, and drug trials start. TIME. 2020 Feb 25. URL: https://time.com/5790545/first-covid-19-vaccine/ [accessed 2021-11-18]
105. New Civil Rights Movement. Trump removes US from global initiative to develop coronavirus treatments and vaccines. Raw Story. 2020 May 02. URL: https://www.rawstory.com/2020/05/trump-removes-us-from-global-initiative-to-develop-coronavirus-treatments-and-vaccines/ [accessed 2021-11-18]
106. Coronavirus: Africa will not be testing ground for vaccine, says WHO. BBC News. 2020 Apr 06. URL: https://www.bbc.com/news/world-africa-52192184 [accessed 2021-11-18]
107. Zeballos-Roig J. Here's why Obamacare would likely make any coronavirus vaccine free for patients — and prove critical in fighting the disease. Business Insider. 2020 Feb 29. URL: https://www.businessinsider.com/coronavirus-vaccine-free-for-patients-result-of-obamacare-2020-2 [accessed 2021-11-18]
108. Florko N. Director of US agency key to vaccine development leaves role suddenly amid coronavirus pandemic. STAT. 2020 Apr 21. URL: https://www.statnews.com/2020/04/21/rick-bright-out-at-barda/ [accessed 2021-11-18]
109. Benkler Y, Faris R, Roberts H. The architecture of our discontent. In: Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. Oxford, UK: Oxford University Press; 2018:45-74.
110. Skogerbø E, Krumsvik AH. Newspapers, Facebook and Twitter: Intermedial agenda setting in local election campaigns. Journal Pract 2014 Sep 02;9(3):350-366. [doi: 10.1080/17512786.2014.950471]

111. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. J Commun Healthc 2020 Dec 15;14(1):12-19. [doi: 10.1080/17538068.2020.1858222]
112. Raghupathi V, Ren J, Raghupathi W. Studying public perception about vaccination: A sentiment analysis of tweets. Int J Environ Res Public Health 2020 May 15;17(10):3464 [FREE Full text] [doi: 10.3390/ijerph17103464] [Medline: 32429223]
113. Simpson E. Proceedings of the 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing. 2018 Presented at: 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing; November 3-7, 2018; Jersey City, NJ p. 237-240. [doi: 10.1145/3272973.3274064]
114. Saxton GD, Niyirora JN, Guo C, Waters RD. #AdvocatingForChange: The strategic use of hashtags in social media advocacy. Adv Soc Work 2015 Jul 27;16(1):154-169. [doi: 10.18060/17952]
115. Broniatowski D, Kerchner D, Farooq F, Huang X, Jamison AM, Dredze M, et al. The COVID-19 social media infodemic reflects uncertainty and state-sponsored propaganda. ArXiv. Preprint posted online on July 19, 2020 [FREE Full text]
116. Walter D, Ophir Y. News frame analysis: An inductive mixed-method computational approach. Commun Methods Meas 2019 Jul 23;13(4):248-266. [doi: 10.1080/19312458.2019.1639145]
117. Kuru O, Stecula D, Lu H, Ophir Y, Chan MS, Winneg K, et al. The effects of scientific messages and narratives about vaccination. PLoS One 2021;16(3):e0248328 [FREE Full text] [doi: 10.1371/journal.pone.0248328] [Medline: 33760856]
118. Cappella JN, Maloney E, Ophir Y, Brennan E. Interventions to correct misinformation about tobacco products. Tob Regul Sci 2015 Jul 01;1(2):186-197. [doi: 10.18001/trs.1.2.8]

Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention
mRNA: messenger RNA
NIH: National Institutes of Health
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.03.21; peer-reviewed by Y Ophir, M Ahmed Kamal, A Majmundar; comments to author 17.04.21; revised version received 11.05.21; accepted 02.10.21; published 03.12.21.

Please cite as:
Cruickshank I, Ginossar T, Sulskis J, Zheleva E, Berger-Wolf T
Content and Dynamics of Websites Shared Over Vaccine-Related Tweets in COVID-19 Conversations: Computational Analysis
J Med Internet Res 2021;23(12):e29127
URL: https://www.jmir.org/2021/12/e29127
doi:10.2196/29127
PMID:34665760

Original Paper

Anti-Asian Sentiments During the COVID-19 Pandemic Across 20 Countries: Analysis of a 12-Billion-Word News Media Database

Reuben Ng[1,2], PhD

[1]Lee Kuan Yew School of Public Policy, National University of Singapore, Singapore, Singapore

[2]Lloyd's Register Foundation Institute for the Public Understanding of Risk, National University of Singapore, Singapore, Singapore

Corresponding Author:
Reuben Ng, PhD
Lee Kuan Yew School of Public Policy
National University of Singapore
469C Bukit Timah Rd
Singapore, 259772
Singapore
Phone: 65 66015229
Email: spprng@nus.edu.sg

Abstract

Background: US president Joe Biden signed an executive action directing federal agencies to combat hate crimes and racism against Asians, which have percolated during the COVID-19 pandemic. This is one of the first known empirical studies to dynamically test whether global societal sentiments toward Asians have become more negative during the COVID-19 pandemic.

Objective: This study aimed to investigate whether global societal sentiments toward Asians across 20 countries have become more negative, month by month, from before the pandemic (October 2019) to May 2020, along with the pandemic (incidence and mortality rates) and cultural (Hofstede's cultural dimensions) predictors of this trend.

Methods: We leveraged a 12-billion-word web-based media database, with over 30 million newspaper and magazine articles taken from over 7000 sites across 20 countries, and identified 6 synonyms of "Asian" that are related to the coronavirus. We compiled their most frequently used descriptors (collocates) from October 2019 to May 2020 across 20 countries, culminating in 85,827 collocates that were rated by 2 independent researchers to provide a Cumulative Asian Sentiment Score (CASS) per month. This allowed us to track significant shifts in societal sentiments toward Asians from a baseline period (October to December 2019) to the onset of the pandemic (January to May 2020). We tested the competing predictors of this trend: pandemic variables of incidence and mortality rates measured monthly for all 20 countries taken from the Oxford COVID-19 Government Response Tracker, and Hofstede's Cultural Dimensions of Individualism, Power Distance, Uncertainty Avoidance, and Masculinity for the 20 countries.

Results: Before the pandemic in December 2019, Jamaica and New Zealand evidenced the most negative societal sentiments toward Asians; when news about the coronavirus was released in January 2020, the United States and Nigeria evidenced the most negative sentiments toward Asians among 20 countries. Globally, sentiments of Asians became more negative—a significant linear decline during the COVID-19 pandemic. CASS trended neutral before the pandemic during the baseline period of October to November 2019 and then plummeted in February 2020. CASS were, ironically, not predicted by COVID-19's incidence and mortality rates, but rather by Hofstede's cultural dimensions: individualism, power distance, and uncertainty avoidance—as shown by mixed models (N=28,494). Specifically, higher power distance, individualism, and uncertainty avoidance were associated with negative societal sentiments toward Asians.

Conclusions: Racism, in the form of Anti-Asian sentiments, are deep-seated, and predicated on structural undercurrents of culture. The COVID-19 pandemic may have indirectly and inadvertently exacerbated societal tendencies for racism. Our study lays the important groundwork to design interventions and policy communications to ameliorate Anti-Asian racism, which are culturally nuanced and contextually appropriate.

(J Med Internet Res 2021;23(12):e28305) doi:10.2196/28305

KEYWORDS

racism; COVID-19; anti-Asian sentiments; psychomics; quantitative social science; culture; text as data; xenophobia; digital humanities

Introduction

The COVID-19 crisis has brought with it a disconcerting increase in racism against the Asian community. In the early phase of the pandemic (March 2020), a Singaporean international student in the United Kingdom was physically assaulted in central London [1]. This was an unprovoked attack that the perpetrators justified by saying "we do not want your coronavirus in our country." Similarly, a Vietnamese curator in London was dropped from an art exhibition, having been told that her presence alone would increase anxiety to those attending the event. She was told that her Asian ethnicity meant that she would be perceived as a carrier of the coronavirus [2].

In the web space, there has been a rise in vitriol and hate speech toward Asians: L1ght found a 900% growth in hate speech towards Asians on Twitter [3]. In the political realm, barbed appellations—"Chinese Virus" and "kung flu" [4]—from government leaders have equated a *whole* ethnicity with the deadly virus and disparaged them, consequently drawing out a fear of Asians. Moreover, the Pew Research Centre found that 58% of Asian Americans reported that it has become more common for people to express racist or racially insensitive views about Asians than before the coronavirus outbreak [5]. Three in 10 Asian American adults have been subject to slurs or jokes because of their race or ethnicity since the outbreak began. In view of the alarming increase in anti-Asian racism, US president Joe Biden signed an executive order on January 26, 2021, directing federal agencies to combat xenophobia toward the Asian American community in the hope of reversing the 2-fold issue of increasing unemployment and hate crimes [6].

Apart from several cross-sectional studies and numerous commentaries [7-10], there are few empirical studies that have assessed the development of anti-Asian sentiments across the pandemic. Our study is one of the first to track anti-Asian sentiments, month by month, from before the pandemic (October 2019) to during the pandemic (May 2020) across 20 countries. We investigated whether global societal sentiments toward Asians across 20 countries have become more negative month by month from before the pandemic (October 2019) to May 2020, as well as the pandemic (incidence and mortality rates) and cultural (Hofstede's cultural dimensions) predictors of this trend.

We selected Hofstede's dimensions [11] as they are one of the most widely used measures of cultural values, and are relevant to this study for 3 reasons. First, they provide multiple dimensions from which a culture can be understood. Second, they are widely recognized and understood, demonstrating our study's comparability and contribution to prior literature. Third, they are based on a national-level understanding of culture, which has been shown to differ from individual-level culture [12], and therefore compatible with the scope of our study as we observe racial sentiments at the national level across 20 countries.

Hofstede's cultural dimensions [13] are as follows: *individualism* is "the degree to which individuals are supposed to look after themselves or remain integrated into groups, usually around the family"; *masculinity* is "the distribution of emotional roles between the genders (…); it opposes 'tough' masculine to 'tender' feminine societies"; *uncertainty avoidance* is "the extent to which a culture programs its members to feel either uncomfortable or comfortable in unstructured situations"; *power distance* is "the extent to which the less powerful members of organizations and institutions accept and expect that power is distributed unequally."

Our study is significant in 3 ways. Conceptually, we contribute to race relations research by analyzing how an extraordinary event, such as the COVID-19 pandemic, may dynamically influence societal sentiments toward an ethnic group. Second, we extend the field of cultural studies by examining the impact of culture on anti-Asian sentiments across 20 countries during a pandemic. Practically, understanding the cultural underpinnings lays the groundwork for designing policy interventions to reduce it, as studies show the malleability of cultural frames [14]—our study provides the cultural considerations to do this effectively.

We hypothesize that societal sentiments toward Asians have become more negative as the pandemic unfolded from October 2019 to May 2020 (hypothesis 1). Second, we tested the factors associated with the hypothesized rise in anti-Asian sentiments during the pandemic. Most opinion-editorials posited that the prevalence of new cases and incidence of new deaths have been associated with increasingly negative racial sentiments (pandemic hypothesis). However, we argue that racism is more deep-seated and is influenced by culture, history, and exclusionary policies such as the Chinese Exclusion Act of 1882 [15]. Discernibly, sociological scholarship reported that cultural factors such as masculine ideals and individualistic traits are associated with racism [16]. Against this background, we hypothesize that both pandemic variables (COVID-19 prevalence and mortality rates) and cultural variables—measured with Hofstede's cultural dimensions—are associated with negative ethnic sentiments during the pandemic (Hypothesis 2).

Methods

Study Design

Data Set

We used the news on the web corpus [17], the largest cross-cultural database with 12 billion words ingested from over 30 million articles—found in over 7000 web-based newspapers and magazines—across 20 countries in six continents: North America (United States and Canada), Asia (Bangladesh, Hong Kong, India, Malaysia, Pakistan, Philippines, Singapore, Sri Lanka), Africa (Ghana, Kenya, Nigeria, South Africa, and Tanzania), the Caribbean islands (Jamaica), Europe (Ireland

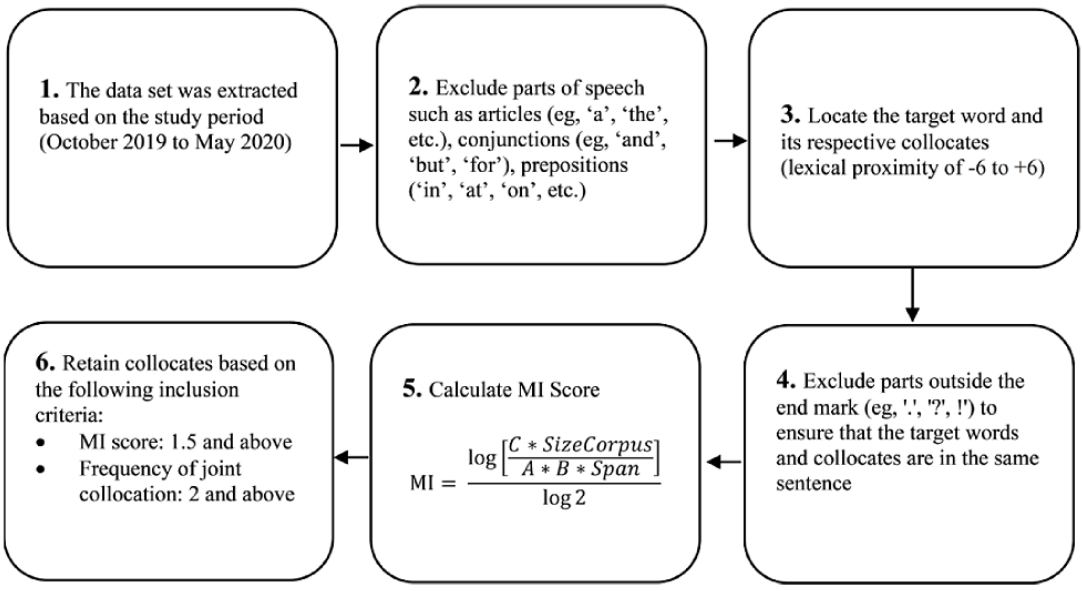
and the United Kingdom), and Oceania (Australia and New Zealand). This data set was created with funding from the National Science Foundation (NSF) and the National Endowment for the Humanities (NEH) to study contemporary language usage in countries where English is mainly and widely used. The dynamic nature of the corpus—with 200 million new words, from 300,000 new articles, added every month—makes it appropriate to study month-by-month change in societal sentiments toward Asians across 8 months, from before the pandemic (October 2019) to during the pandemic (May 2020). Within this study's timeframe, our corpus consists of 1.5 billion words. This data set is appropriate as Cultivation Theory [18] suggests that the large representation of web-based media within the corpus reflects societal perceptions across 20 countries and provides an extraordinary platform to study the global sentiments of Asians. This unprecedented data set further circumvents the ecological fallacy that limits most survey studies that mistakenly draw country-level conclusions from individual-level surveys.

Data Preprocessing

The data set was preprocessed through a 6-step process that aimed at identifying a list of 7 relevant target nouns (*Asian*, *Asians*, *Asia*, *Chinese*, *China*, *Wuhan*, and *Hubei*) and their descriptors (collocates) for sentiment analysis. Figure 1 shows a flowchart for data preprocessing. We generated collocates (ie, words that co-occurred most frequently with each target noun) every month, between October 2019 and May 2020 for each of the 20 countries with the following inclusion criteria: (1) lexical proximity: collocates present within 6 words prior and after the target noun (step 3, Figure 1). Articles such as "the" and "a" were not included in the 6-word lexical span. If the target noun was the first word of a sentence, the collocates from the prior sentence were excluded. (2) Mutual Information (MI) score of 3 and above: collocates had a stronger association with the target noun than with other nouns in the corpus for that country, indicating semantic bonding (step 4, Figure 1) [19]. The MI score estimates word association norms directly from the corpus. It is calculated via sentiment analysis, which shows the MI between collocates and target words. The higher the MI value, the closer the relationship between the collocate and target word. The MI value is calculated using the formula:

where *A* indicates the possibility of the target word *A* appearing, which is calculated by the frequency of the target word. *B* indicates the possibility of the collocate *B* appearing, which is calculated by the frequency of word *B*. *C* indicates the possibility of *A* and *B* appearing together, which is calculated by the frequency of collocate *B* appearing near the target word *A*. "SizeCorpus" refers to the size of corpus or the number of words. Span is the span of words (eg, if there are 6 words to the left and 6 words to the right of the target word, span=12) [log (2)=0.30103]. This is a well-established application of computational linguistics to study stereotypes in other studies [20-28]. The rigorous process culminated in 85,827 collocates over eight months, across 20 countries.

Figure 1. Preprocessing steps to identify collocates of target nouns for sentiment analysis. MI: Mutual Information.



Measurement of Societal Sentiments of Asians Across 20 Countries

After data preprocessing, each collocate was rated on a scale from 1 (very negative) to 5 (very positive) by 2 independent raters with a high interrater reliability (Cronbach α) of .916 (95% CI 0.874-0.958). For instance, *dirty*, *dangerous*, and *suspicious* were rated 1 (very negative); *employee*, *resident*, and *transportation* were rated 3 (neutral); *empowering*, *hero*, and *venerable* were rated 5 (very positive). Table 1 provides sample descriptors associated with narratives on Asians. We calculated a Cumulative Asian Sentiment Score (CASS) per month by taking the respective mean ratings globally to test Hypothesis 1, and by country, to test hypothesis 2. CASS is mathematically defined as follows:

where Y_{kt} is the CASS for country K during month t. For country k, during month t, S_i is the score of the word (i), while h_{ikt} is the count of the word (i).

Table 1. Examples of negative, neutral, and positive descriptors of Asians in the media during the COVID-19 pandemic.

Negative	Neutral	Positive
sin	commerce	benefit
weakness	secretary	leading
vile	domestic	calm
protest	affair	rejuvenated
vilify	demand	intelligence
denounce	military	good
deface	philosopher	peaceful
anti-china	deal	intellectual
anger	navy	groundbreaking
punishment	culture	successfully
crisis	market	sophisticated
chaos	country	greatest
agitation	establish	charm
meddling	influence	admiration
discord	rule	improving

Pandemic Variables

The monthly prevalence of new cases and incidences of new deaths across 20 countries were derived from the Oxford COVID-19 Government Response Tracker [29].

Hofstede's Cultural Dimensions

Calculations of the country dimension scores have been reported by Hofstede and Minkov [30], which were based on original cross-national surveys of IBM employees, and subsequent studies (eg, Hofstede Insights, 2020). The country score for each of Hofstede's dimensions was calculated using this method: First, individual survey responses to each question were calculated at the national level. For questions that were answered on a 5-point Likert scale, the national mean of the responses was calculated. For questions requiring a yes/no or multiple-choice answer, the national percentage of each answer, or a combination of answers (eg, "Option A OR Option C") was calculated. Next, these national-level scores were combined using a weighted formula to yield a country dimension score based on 3-8 survey questions, resulting in final scores that ranged from 0 to 100.

Analytical Strategy

Hypothesis 1, positing that societal sentiments toward Asians (as measured with the CASS) have become more negative across 8 months from October 2019 to May 2020, was tested using a trend model to show a significant linear decline in racial sentiments. The linear regression model can be expressed as the following:

$$y = a + bX$$

Where a represents the intercept, b represents the slope and X represents the week.

Hypothesis 2, positing that pandemic variables (COVID-19 incidence and mortality rates) and cultural variables (Hofstede's cultural dimensions) are associated with the CASS, was tested using a linear mixed model. CASS was the independent variable while the pandemic and cultural variables were the dependent variables. The model can be expressed as the following:

$$y = X\beta + Zb + \varepsilon$$

Where $b \sim N(0, \varphi_\theta)$ and $\sim N(0, \gamma_\theta)$, X and Z represent model matrices for the fixed effects—pandemic and cultural dimensions—(β) and random effects—time—(b), respectively. All data preprocessing, text analytics, and statistical analyses were conducted using Python (version 3.7) and OriginPro 2019b.

Data Accessibility

Data are publicly available at English-Corpora.org.

Results

CASS Across 20 Countries During the COVID-19 Pandemic

Globally, we observed a significant negative linear trend in CASS toward Asians (β=–.03, P=.018), supporting hypothesis 1 (Figure 2). Other trends such as quadratic and cubic were nonsignificant. Societal sentiments of Asians, measured with the CASS, were trending neutral before the pandemic (October to December 2019) and reached the lowest point in February 2020 during the pandemic. Before the pandemic in December 2019, Jamaica and New Zealand evidenced the most negative

societal sentiments toward Asians. In January 2020, when news of the novel coronavirus was released, the United States and Nigeria evidenced the most negative societal sentiments of Asians among 20 countries. The CASS for each of the 20 countries are presented in Figures 3 and 4.

Figure 2. Global Cumulative Asian Sentiment Scores (CASS) across 20 countries from October 2019 to May 2020. Globally, societal sentiments of Asians became more negative across 20 countries—a significant linear decline during the COVID-19 pandemic. CASS were trending neutral before the pandemic (October to November 2019), and plummeted in February 2020. Thereafter, CASS increased gradually, although it has not recovered to prepandemic levels.

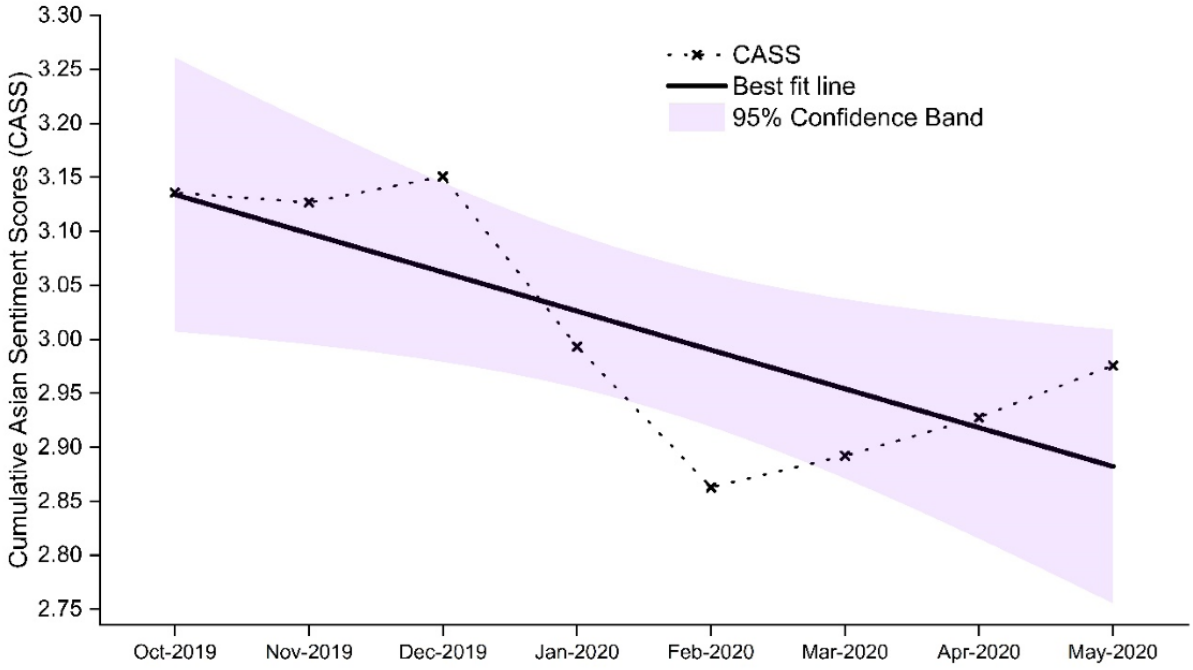


Figure 3. Cumulative Asian Sentiment Scores (CASS) across 10 countries (Australia, Bangladesh, Ghana, Hong Kong, Jamaica, Kenya, Nigeria, New Zealand, Singapore, and Tanzania) from October 2019 to May 2020. There was a significant negative linear trend in CASS across the various countries as the pandemic progressed.

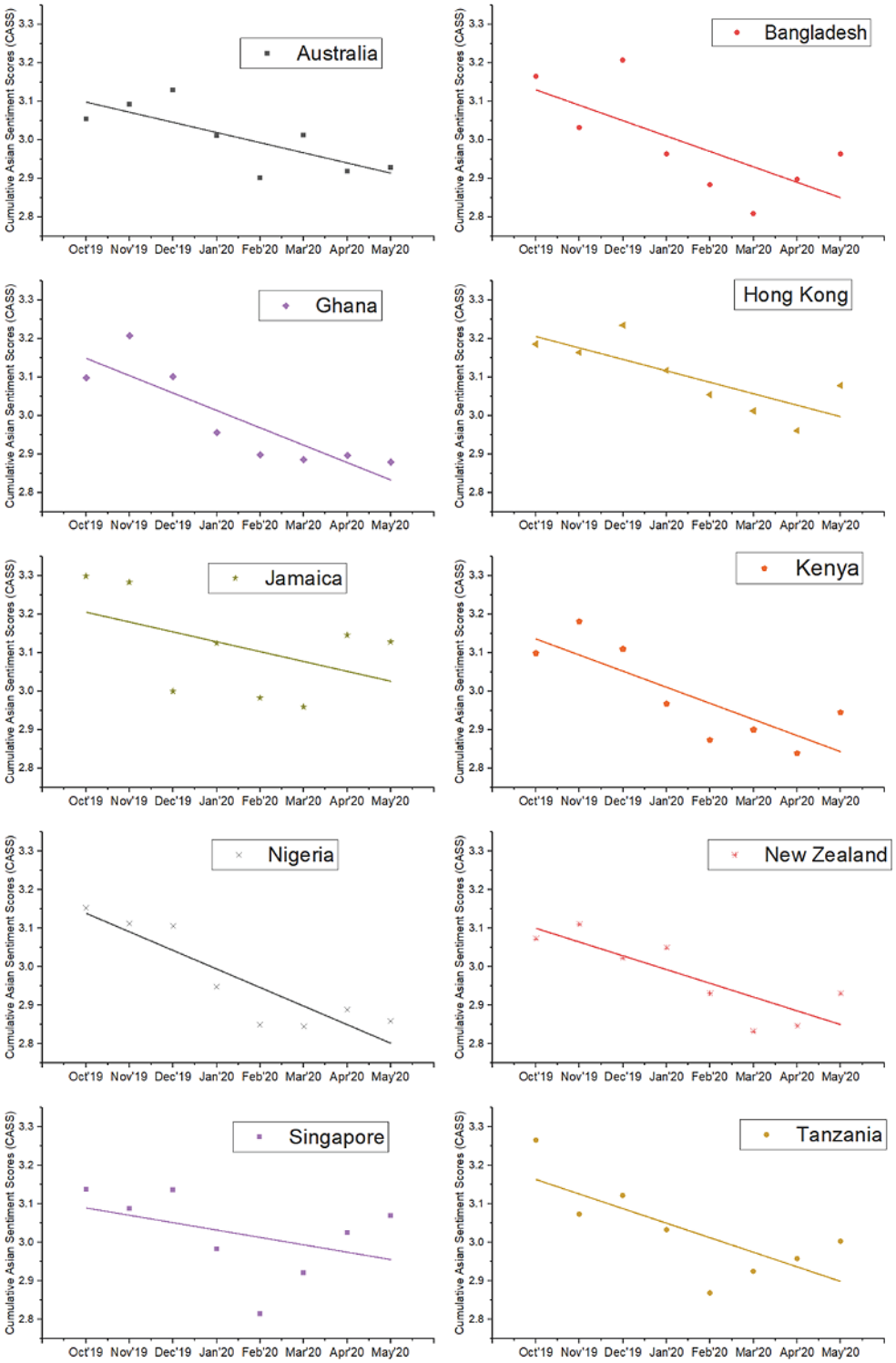
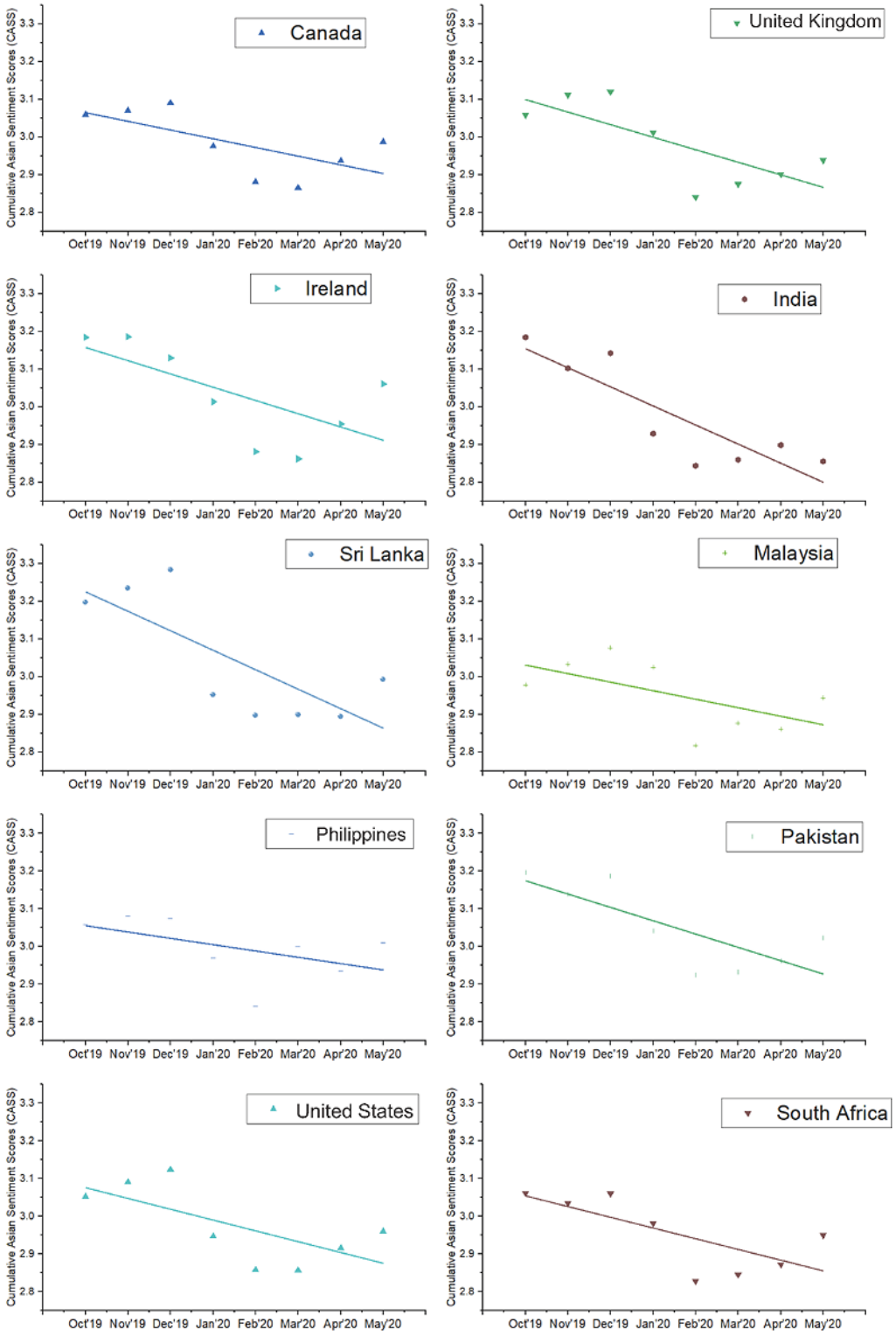


Figure 4. Cumulative Asian Sentiment Scores (CASS) across 10 countries (Canada, the United Kingdom, Ireland, India, Sri Lanka, Malaysia, Philippines, Pakistan, the United States, and South Africa) from October 2019 to May 2020. There was a significant negative linear trend in CASS across the various countries as the pandemic progressed.



Predictors of Racial Sentiment Trends During the COVID-19 Pandemic

We tested hypothesis 2 using a mixed model, with pandemic variables (monthly prevalence of new cases and incidence of new deaths for 20 countries) and cultural dimensions as fixed variables, and time as the random variable (Table 2). The pandemic variables of prevalence and mortality rates did not reach significance, while 3 of 4 cultural dimensions were significant: higher power distance (β=–.002, *P*<.001), individualism (β=–.002, *P*<.001), and uncertainty avoidance (β=–.001, *P*<.001) were associated with negativity of societal sentiments toward Asians, holding other covariates constant; this provides support for hypothesis 2.

Table 2. Pandemic and cultural predictors of societal sentiments toward Asians across 20 countries during the COVID-19 pandemic (N=28,494): Model 1.[a] Higher power distance, individualism and uncertainty avoidance were associated with negative sentiments toward Asians; the pandemic variables of prevalence and mortality rates did not reach statistical significance.

Predictor	Value (SE)
Prevalence of new cases[b]	0.000 (0.000)
Incidence of new deaths[c]	0.000 (0.000)
Individualism	–0.002[d] (0.000)
Power distance	–0.002[d] (0.000)
Uncertainty avoidance	–0.001[d] (0.000)
Masculinity	–0.001 (0.000)

[a]R^2=0.737; adjusted R^2=0.723.

[b]Respective monthly COVID-19 incidence rates from October 2019 to May 2020.

[c]Respective monthly COVID-19 mortality rates from October 2019 to May 2020.

[d]P<.001.

Discussion

This is one of the first known empirical studies to dynamically test whether global societal sentiments toward Asians have become more negative month by month from before the pandemic (October 2019) to during the pandemic (May 2020). Our findings show that societal sentiments toward Asians became more negative across the 20 countries over time. This study contributes to the existing literature in 2 ways. First, it contributes to research on race relations by showing how pandemics may have severe repercussions on discourse surrounding certain ethnic groups. Second, by providing insight into the cultural factors associated with anti-Asian sentiments, this study lays the groundwork for more effective cultural communication strategies concerning public health issues.

Historically, incidences of xenophobia have accompanied the onset of global pandemics. With such catastrophic outbreaks, fear often drives those at risk to pin the blame on some group external to their own national, religious, or ethnic identity—in a bid to attribute their troubles to some known entity while reinforcing an “us against them” mentality [31]. As sickness cultivates fear, it can in turn incite moral panic and promote bias. Consequently, minority groups have often found themselves erroneously blamed for being contagious, as others perceive them to be “dirty” [32]. This blame culture could be a driving force for the negativity of societal sentiments toward Asians during the pandemic, as shown in our study.

Delving further, we argue that the negative rhetoric against Asians promoted “otherness” [32-34]. In times of crises such as the current pandemic, it is common for individuals with similar attributes to be categorized as the ingroup, and those who are “different” as the outgroup. The differentiation accentuates the similarities and homogeneity of ingroup members, while juxtaposing the differences of the outgroup members who have been accused of spreading the virus. This reinforces ingroup notions of “normality,” and positions those who are different—the Asians in this context—as deviant, through a process of othering and social exclusion [34]. A consequence of this othering could be prejudice, unfavorable treatment, and discrimination against Asians—possibly leading to heightened racial tensions and frayed social fabric in multiracial communities.

In addition, metaphors are ubiquitous in forming our world view [35]. The adjective “Chinese” in “Chinese virus” is particularly problematic as it metaphorizes the devastation of the pandemic with an ethnicity. Connecting group identities with explicitly negative medical language (ie, virus) serves to categorize those group identities as others. Such “othering” rhetoric only serves to typify stigmatized groups, devaluing them as those in society who possess undesirable characteristics that are outside normal expectations [36].

In relation to hypothesis 2, our results show how anti-Asian sentiments during the COVID-19 pandemic were strongly associated with cultural norms in society and not with the pandemic variables. Three of 4 cultural dimensions [37-42]—individualism, power distance, and uncertainty avoidance—were associated with increasing negativity of racial sentiments. This is one of the first known studies to link cultural dimensions and race relations during the COVID-19 pandemic, and this study attempted to explain these links.

First, we found that increased individualism was associated with negative societal sentiments toward Asians. In societies where tenets of individualism thrive, self-interests and outward expressions of intense emotions such as anger and antagonism [11] are more common, with less importance placed on avoiding confrontation and social harmony. We postulate that high levels of individualism in some societies could have encouraged people to speak their minds and outwardly display of anger and blame toward Asians, resulting in an outpouring of more pernicious sentiments. Moreover, in high-individualism societies, epitomized by a desire for an autonomous identity defined by personal choices and achievements [11], feelings of intrusion and encroachment into these pursuits—inflicted by the virus and those deemed to be “associated” with it—might have also ignited the spark for such confrontational anti-Asian sentiments. We argue that the widespread fear of contracting the disease and thus being precluded from pursuing one’s goals freely engenders prejudice against the outgroup that might be deemed

responsible for this, positioning them as effective scapegoats for both the virus and hampering one's personal individualistic identity.

Second, high power distance was associated with negative societal sentiments toward Asians. In essence, power distance represents the amount of respect and deference between people from different levels of the social hierarchy. Cultures high in power distance advocate a social hierarchy based on institutionalized status differentials, where members from the lower status groups are disadvantaged in terms of resource allocation, are expected to treat their high-status counterparts with deference and respect [11]. We argue that such demarcations of inequality allow for a perpetuation of social dichotomies and an increase in xenophobic sentiments toward those deemed undeserving of respect owing to their perceived "lower" status.

Third, high uncertainty avoidance was associated with negative societal sentiments toward Asians. We suggest that the uncomfortable threat and dangers from stemming those "associated" with the virus, and the pressing need to actively dispel the anxiety of its looming impact, has led to the prevalence of such negative narratives toward Asians. Labeling others on the basis of a variety of stigmas and associating an entire race with the virus not only aids in the identification of "negative" traits that are synonymous with risk, but also reinforces the lack of tolerance for ambiguities and uncertainty associated with particular groups of people who have been deemed to possess deviant behaviors [43]. Such stigmatizing sentiments serve as the cognitive basis of social grouping, helping individuals categorize and cautiously avoid people of particular social groups by displacing the threat of the unknown onto them. We thus argue that uncertainty avoidance is linked to stereotyping, distrusting, and dehumanizing outgroups through harsh sentiments owing to the perceived unpredictability of outgroup members possibly "spreading" the disease and jeopardizing them.

Hofstede's cultural dimensions have proven to be a useful framework to explore how anti-Asian sentiments have evolved over the course of the COVID-19 pandemic. It is evident that deep-seated cultural norms—individualism, power distance, and uncertainty avoidance—are a driving force in stigmatizing behavior against outgroups and revealed apertures in societal relations that arise during global crises such the COVID-19 pandemic. Future studies should explore the associations between cultural dimensions and racism. In addition, our study limitations could provide ideas for future studies. A key shortcoming is the lack of data from social media in the corpus. The diversity of social media usage across multiple platforms renders data collation challenging, and most social media platforms such as Facebook are closed for public access—they have also become increasingly monetized, selling selected data sets that may not be representative. Though our methodology allows for dynamic social sensing, compared to traditional methods [44-54], this is a significant drawback that we intend to overcome in future studies upon augmenting our database.

The COVID-19 pandemic has deepened fissures in racial relations as posited by numerous commentaries, and now shown empirically by the negativity of societal sentiments of Asians, across 20 countries, over 8 months, from October 2019 to May 2020. It is critical that governments and policy makers consider the cultural underpinnings linked to anti-Asian sentiments when designing appropriate messages for risk communication. In particular, societies high in power distance, individualism and uncertainty avoidance must be prioritized. Confronting the specific cultural factors fueling anti-Asian sentiment will achieve a multiplier effect, eventually forming the basis for solidarity.

Acknowledgments

We gratefully acknowledge support from the Social Science Research Council SSHR Fellowship (MOE2018-SSHR-004), and the Lloyd's Register Foundation IPUR Grant (IPUR-FY2019-RES-03-NG). We are also indebted to WY, NI, and SJ for research assistance. The funders had no role in study design, data collection, analysis, writing, and decision to publish.

Authors' Contributions

RN designed the study, developed the methodology, analyzed the data, led the writing of the manuscript, and acquired the research funding.

Conflicts of Interest

None declared.

References

1. Cheng SO. Xenophobia due to the coronavirus outbreak - A letter to the editor in response to "the socio-economic implications of the coronavirus pandemic (COVID-19): A review". Int J Surg 2020 Jul;79:13-14 [FREE Full text] [doi: 10.1016/j.ijsu.2020.05.017] [Medline: 32407798]
2. Coates M. Covid-19 and the rise of racism. BMJ 2020 Apr 06;369:m1384. [doi: 10.1136/bmj.m1384] [Medline: 32253178]
3. L1GHT. Rising Levels of Hate Speech & Online Toxicity During This Time of Crisis. L1GHT. URL: https://l1ght.com/Toxicity_during_coronavirus_Report-L1ght.pdf [accessed 2021-02-01]

4. Boyer D. Trump spars with reporter over accusation that staffer called coronavirus ‘Kung flu’. The Washington Times. 2020 Mar 18. URL: https://www.washingtontimes.com/news/2020/mar/18/trump-spars-reporter-over-accusation-staffer-calle/ [accessed 2021-02-01]
5. Ruiz N, Horowitz J, Tamir C. Many Black, Asian Americans Say They Have Experienced Discrimination Amid Coronavirus. Pew Research Center. 2020 Jul 01. URL: https://www.pewsocialtrends.org/2020/07/01/many-black-and-asian-americans-say-they-have-experienced-discrimination-amid-the-covid-19-outbreak/#three-in-ten-or-more-u-s-adults-say-racist-views-about-asian-and-black-americans-are-more-common-than-before-the-pandemic [accessed 2021-02-01]
6. Memorandum Condemning and Combating Racism, Xenophobia, and Intolerance Against Asian Americans and Pacific Islanders in the United States. The White House. 2021 Jan 26. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/26/memorandum-condemning-and-combating-racism-xenophobia-and-intolerance-against-asian-americans-and-pacific-islanders-in-the-united-states/ [accessed 2021-02-01]
7. Chung RY, Li MM. Anti-Chinese sentiment during the 2019-nCoV outbreak. Lancet 2020 Feb 29;395(10225):686-687 [FREE Full text] [doi: 10.1016/S0140-6736(20)30358-5] [Medline: 32122469]
8. Roberto KJ, Johnson AF, Rauhaus BM. Stigmatization and prejudice during the COVID-19 pandemic. Adm Theory Prax 2020 Jun 25;42(3):364-378. [doi: 10.1080/10841806.2020.1782128]
9. 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]
10. Whitehead D. 'You deserve the coronavirus': Chinese people in UK abused over outbreak. Sky News. 2020 Feb 12. URL: https://news.sky.com/story/coronavirus-chinese-people-face-abuse-in-the-street-over-outbreak-11931779 [accessed 2021-02-01]
11. Hofstede G. Culture’s Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Thousand Oaks, CA: Sage Publications; 2001.
12. Zhang X, Xing C, Guan Y, Song X, Melloy R, Wang F, et al. Attitudes toward older adults: A matter of cultural values or personal values? Psychol Aging 2016 Mar;31(1):89-100. [doi: 10.1037/pag0000068] [Medline: 26691301]
13. Hofstede G. Dimensionalizing Cultures: The Hofstede Model in Context. ORPC 2011 Dec 01;2(1). [doi: 10.9707/2307-0919.1014]
14. Dweck CS, Chiu C, Hong Y. Implicit Theories and Their Role in Judgments and Reactions: A Word From Two Perspectives. Psychol Inq 2009 Nov 19;6(4):267-285. [doi: 10.1207/s15327965pli0604_1]
15. Yung J, Chang G, Lai H. Chinese American Voices: From the Gold Rush to the Present. Berkeley, CA: University of California Press; 2012.
16. Dovidio J, Gaertner S. Prejudice, Discrimination, and Racism. San Diego, CA: Academic Press; 1992.
17. Davies M. The new 4.3 billion word NOW corpus, with 4--5 million words of data added every day. 2017. URL: https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2017/general/paper250.pdf [accessed 2021-10-29]
18. Gerbner G, Gross L, Morgan M, Signorielli N, Shanahan J. Growing up with television: Cultivation processes. In: Bryant J, Zillmann D, editors. Media effects: Advances in theory and research. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2002:43-67.
19. Church K, Hanks P. Word association norms, mutual information, and lexicography. In: Proceedings of the 27th annual meeting on Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 1989:26-83.
20. Ng R. Societal Age Stereotypes in the U.S. and U.K. from a Media Database of 1.1 Billion Words. Int J Environ Res Public Health 2021 Aug 21;18(16):8822 [FREE Full text] [doi: 10.3390/ijerph18168822] [Medline: 34444578]
21. Ng R, Allore HG, Trentalange M, Monin JK, Levy BR. Increasing negativity of age stereotypes across 200 years: evidence from a database of 400 million words. PLoS One 2015;10(2):e0117086 [FREE Full text] [doi: 10.1371/journal.pone.0117086] [Medline: 25675438]
22. Ng R, Lim W. Ageism linked to culture, not demographics: Evidence from an 8-billion-word corpus across 20 countries. J Gerontol B Psychol Sci Soc Sci 2020 Oct 25;76(9):1791-1798. [doi: 10.1093/geronb/gbaa181] [Medline: 33099600]
23. Ng R, Chow T. Aging Narratives over 210 years (1810-2019). J Gerontol B Psychol Sci Soc Sci 2020 Dec 10;76(9):1799-1807 [FREE Full text] [doi: 10.1093/geronb/gbaa222] [Medline: 33300996]
24. Ng R, Chow T, Yang W. Culture Linked to Increasing Ageism during Covid-19: Evidence from a 10-billion-word Corpus across 20 Countries. J Gerontol B Psychol Sci Soc Sci 2021 Mar 31;76(9):1808-1816 [FREE Full text] [doi: 10.1093/geronb/gbab057] [Medline: 33786581]
25. Ng R, Chow T, Yang W. News media narratives of Covid-19 across 20 countries: Early global convergence and later regional divergence. PLoS One 2021;16(9):e0256358 [FREE Full text] [doi: 10.1371/journal.pone.0256358] [Medline: 34469446]

26. Ng R, Indran N. Societal perceptions of caregivers linked to culture across 20 countries: Evidence from a 10-billion-word database. PLoS One 2021;16(7):e0251161 [FREE Full text] [doi: 10.1371/journal.pone.0251161] [Medline: 34197470]
27. Ng R, Indran N. Role-Based Framing of Older Adults Linked to Decreased Ageism Over 210 Years: Evidence From a 600-Million-Word Historical Corpus. Gerontologist 2021 Jul 29:gnab108. [doi: 10.1093/geront/gnab108] [Medline: 34323967]
28. Ng R, Indran N. Reframing Aging during COVID-19: Familial Role-Based Framing of Older Adults Linked to Decreased Ageism. J Am Geriatr Soc 2021 Oct 21. [doi: 10.1111/jgs.17532] [Medline: 34674224]
29. COVID-19 Government Response Tracker. Blavatnik School of Government, University of Oxford. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker [accessed 2021-02-01]
30. Hofstede G, Minkov M. VSM 2013: Values Survey Module 2013 Manual. 2013. URL: https://www.laits.utexas.edu/orkelm/kelmpub/VSM2013_Manual.pdf [accessed 2021-10-29]
31. Ng R, Indran N. Societal Narratives on Caregivers in Asia. IJERPH 2021 Oct 26;18(21):11241. [doi: 10.3390/ijerph182111241]
32. Muzzatti SL. Bits of Falling Sky and Global Pandemics: Moral Panic and Severe Acute Respiratory Syndrome (SARS). Illness, Crisis & Loss 2017 Dec 04;13(2):117-128. [doi: 10.1177/105413730501300203]
33. Ng R, Tan Y. Diversity of COVID-19 News Media Coverage across 17 Countries: The Influence of Cultural Values, Government Stringency and Pandemic Severity. Int J Environ Res Public Health 2021.
34. Tajfel H, Turner J. The Social Identity Theory of Intergroup Behavior. In: Key Readings in Social Psychology. Hove: Psychology Press; 2004:76.
35. Lakoff G, Johnson M. Metaphors We Live By. Chicago, IL: University of Chicago Press; 1981.
36. Goffman E. Stigma Notes on the Management of Spoiled Identity. New York, NY: Penguin Books; 1990.
37. Ng R. Douglas, Mary. In: The Encyclopedia of Cross-Cultural Psychology. Hoboken, NJ: John Wiley & Sons; 2013:439.
38. Ng R. Cultural Theory. In: The Encyclopedia of Cross-Cultural Psychology. Hoboken, NJ: John Wiley & Sons; 2013:333.
39. Ng R. Cultural Intelligence. In: The Encyclopedia of Cross-Cultural Psychology. Hoboken, NJ: John Wiley & Sons; 2013:310.
40. Ng R. Cultural Cognition. In: The Encyclopedia of Cross-Cultural Psychology. Hoboken, NJ: John Wiley & Sons; 2013:300.
41. Ng R, Rayner S. Integrating psychometric and cultural theory approaches to formulate an alternative measure of risk perception. Innovation Eur J Social Sci Res 2010 Jun;23(2):85-100 [FREE Full text] [doi: 10.1080/13511610.2010.512439]
42. Steglitz J, Ng R, Mosha J, Kershaw T. Divinity and distress: the impact of religion and spirituality on the mental health of HIV-positive adults in Tanzania. AIDS Behav 2012 Nov;16(8):2392-2398. [doi: 10.1007/s10461-012-0261-7] [Medline: 22797930]
43. Gudykunst W, Ting-Toomey S, Nishida T. Communication in Personal Relationships Across Cultures. Thousand Oaks, CA: Sage Publications; 1996.
44. Heidekrueger P, Juran S, Szpalski C, Larcher L, Ng R, Broer P. The current preferred female lip ratio. J Craniomaxillofac Surg 2017 May;45(5):655-660. [doi: 10.1016/j.jcms.2017.01.038] [Medline: 28318919]
45. Sima L, Ng R, Elimelech M. Modeling risk categories to predict the longitudinal prevalence of childhood diarrhea in Indonesia. Am J Trop Med Hyg 2013 Nov;89(5):884-891 [FREE Full text] [doi: 10.4269/ajtmh.12-0540] [Medline: 24019442]
46. Ng R. Cloud Computing in Singapore: Key Drivers and Recommendations for a Smart Nation. PaG 2018 Nov 21;6(4):39-47 [FREE Full text] [doi: 10.17645/pag.v6i4.1757]
47. Ng R, Lim S, Saw S, Tan K. 40-Year Projections of Disability and Social Isolation of Older Adults for Long-Range Policy Planning in Singapore. Int J Environ Res Public Health 2020 Jul 09;17(14):4950 [FREE Full text] [doi: 10.3390/ijerph17144950] [Medline: 32659983]
48. Ng R, Levy B. Pettiness: Conceptualization, measurement and cross-cultural differences. PLoS One 2018;13(1):e0191252 [FREE Full text] [doi: 10.1371/journal.pone.0191252] [Medline: 29385157]
49. Ng R, Tan K. Implementing an Individual-Centric Discharge Process across Singapore Public Hospitals. Int J Environ Res Public Health 2021 Aug 17;18(16):8700 [FREE Full text] [doi: 10.3390/ijerph18168700] [Medline: 34444448]
50. Ng R, Allore H, Monin J, Levy B. Retirement as Meaningful: Positive Retirement Stereotypes Associated with Longevity. J Soc Issues 2016 Mar;72(1):69-85 [FREE Full text] [doi: 10.1111/josi.12156] [Medline: 27346893]
51. Ng R, Ang R, Ho M. Coping with Anxiety, Depression, Anger and Aggression: The Mediational Role of Resilience in Adolescents. Child Youth Care Forum 2012 May 13;41(6):529-546 [FREE Full text] [doi: 10.1007/s10566-012-9182-x]
52. Ng R, Wong PSH. Marginalization of Graduate Freelancers in the Gig Economy. Japan Labor Issues 2021;5(32):67-84 [FREE Full text]
53. Ng R, Allore H, Levy B. Self-Acceptance and Interdependence Promote Longevity: Evidence From a 20-year Prospective Cohort Study. Int J Environ Res Public Health 2020 Aug 18;17(16):5980 [FREE Full text] [doi: 10.3390/ijerph17165980] [Medline: 32824658]
54. Ng R, Chow T, Yang W. The Impact of Aging Policy on Societal Age Stereotypes and Ageism. Gerontologist 2021 Oct 12:gnab151. [doi: 10.1093/geront/gnab151] [Medline: 34636402]

Abbreviations

CASS: Cumulative Asian Sentiment Score
MI: Mutual Information
NEH: National Endowment for the Humanities
NSF: National Science Foundation

Edited by G Eysenbach; submitted 28.02.21; peer-reviewed by M Paggi, H Lee; comments to author 23.06.21; revised version received 28.07.21; accepted 03.10.21; published 08.12.21.

Please cite as:
Ng R
Anti-Asian Sentiments During the COVID-19 Pandemic Across 20 Countries: Analysis of a 12-Billion-Word News Media Database
J Med Internet Res 2021;23(12):e28305
URL: https://www.jmir.org/2021/12/e28305
doi:10.2196/28305
PMID:34678754

Original Paper

Identifying Insomnia From Social Media Posts: Psycholinguistic Analyses of User Tweets

Ahmed Shahriar Sakib[1], BSc; Md Saddam Hossain Mukta[2], PhD; Fariha Rowshan Huda[1], BSc; A K M Najmul Islam[3], PhD; Tohedul Islam[1], MSc; Mohammed Eunus Ali[4], PhD

[1]American International University-Bangladesh, Dhaka, Bangladesh

[2]United International University, Dhaka, Bangladesh

[3]LUT University, Lappeenranta, Finland

[4]Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

Corresponding Author:
Md Saddam Hossain Mukta, PhD
United International University
Madani Ave
Natun Bazar
Dhaka, 1216
Bangladesh
Phone: 880 1712 095216
Email: saddam@cse.uiu.ac.bd

Abstract

Background: Many people suffer from insomnia, a sleep disorder characterized by difficulty falling and staying asleep during the night. As social media have become a ubiquitous platform to share users' thoughts, opinions, activities, and preferences with their friends and acquaintances, the shared content across these platforms can be used to diagnose different health problems, including insomnia. Only a few recent studies have examined the prediction of insomnia from Twitter data, and we found research gaps in predicting insomnia from word usage patterns and correlations between users' insomnia and their Big 5 personality traits as derived from social media interactions.

Objective: The purpose of this study is to build an insomnia prediction model from users' psycholinguistic patterns, including the elements of word usage, semantics, and their Big 5 personality traits as derived from tweets.

Methods: In this paper, we exploited both psycholinguistic and personality traits derived from tweets to identify insomnia patients. First, we built psycholinguistic profiles of the users from their word choices and the semantic relationships between the words of their tweets. We then determined the relationship between a users' personality traits and insomnia. Finally, we built a double-weighted ensemble classification model to predict insomnia from both psycholinguistic and personality traits as derived from user tweets.

Results: Our classification model showed strong prediction potential (78.8%) to predict insomnia from tweets. As insomniacs are generally ill-tempered and feel more stress and mental exhaustion, we observed significant correlations of certain word usage patterns among them. They tend to use negative words (eg, "no," "not," "never"). Some people frequently use swear words (eg, "damn," "piss," "fuck") with strong temperament. They also use anxious (eg, "worried," "fearful," "nervous") and sad (eg, "crying," "grief," "sad") words in their tweets. We also found that the users with high neuroticism and conscientiousness scores for the Big 5 personality traits likely have strong correlations with insomnia. Additionally, we observed that users with high conscientiousness scores have strong correlations with insomnia patterns, while negative correlation between extraversion and insomnia was also found.

Conclusions: Our model can help predict insomnia from users' social media interactions. Thus, incorporating our model into a software system can help family members detect insomnia problems in individuals before they become worse. The software system can also help doctors to diagnose possible insomnia in patients.

(J Med Internet Res 2021;23(12):e27613) doi:10.2196/27613

KEYWORDS

insomnia; Twitter; word embedding; Big 5 personality traits; classification; social media; prediction model; psycholinguistics

Introduction

Background

Insomnia, a type of sleep disorder, is the inability to fall asleep or stay asleep at night. It is one of the most prevalent mental health symptoms globally [1]. One study [2] suggests that approximately 30% of adults worldwide exhibit insomnia symptoms, like difficulty initiating and maintaining sleep and waking up too early. People with insomnia might also experience other problems, such as depression, anxiety, and excessive alcohol consumption [3].

With the unprecedented growth of smartphone and internet technologies, social media has now become a ubiquitous platform that reflects users' daily activities, preferences, and beliefs. These social media platforms have become a means to share health information [4-9] for many users. For example, Paul et al [10] have stated that Twitter has become a common place to discuss a wide range of health information, including insomnia and other mental health conditions such as depression, stress, and anxiety. Several other studies [10-12] also report that Twitter is used as a platform to share symptoms [11], seek help, and exchange advice [12].

As insomnia is a mental health disorder, the illness might have a strong connection with human personality attributes. In fact, prior studies [13-15] suggest that insomnia has links with certain personality traits. Therefore, in this study, we attempted to derive personality traits from users' social media interactions and use these traits along with users' word usage patterns to predict insomnia. To the best of our knowledge, our study is the first to investigate from social media interactions whether personality traits have an association with insomnia. Predicting insomnia by analyzing users' tweets has a number of real-life applications. For example, friends and parents can identify problems in their loved ones, while health care providers can use the system to diagnose insomnia and can build an automated early warning system.

Insomnia and the Big 5 Personality Traits

Many adults experience short-term (acute) insomnia, which lasts for days or weeks. Acute insomnia is common and often brought on by situations such as stress at work, family pressures, or a traumatic event. Some adults have long-term (chronic) insomnia that lasts for several months or years [16]. In most cases, chronic insomnia may be a side effect of other problems [16]. Insomnia not only reduces individuals' energy levels but also degrades their health, work performance, and quality of life. There are several causes of insomnia [16-18], including mental health disorders, such as posttraumatic stress disorder. Antidepressants, asthma medications, and blood pressure medications can also lead to sleep disorders. Medical conditions such as chronic pain, cancer, diabetes, heart disease, asthma, gastroesophageal reflux disease, overactive thyroid, Parkinson disease, and Alzheimer disease can also cause insomnia. High consumption of caffeine, nicotine, or alcohol may also prevent sleep and lead to insomnia [16].

Personality differentiates individuals in their patterns of thinking, feeling, and behaving [19]. The Big 5 personality scale is a common scale for measuring personality [19,20]. The Big 5 model has 5 different personality traits: openness, conscientiousness, extroversion, agreeableness, and neuroticism. People with a high level of openness have a tendency to reflect on ideas, innovate, and appreciate values. People with high conscientiousness are cautious and meticulous and tend to seek achievement. People with high extroversion have a tendency to seek excitement and show positive emotions. People with high agreeableness tend to be sympathetic, trusted, and merciful to others. Neurotic individuals show negative emotions, such as anxiety, inhibition, anger, and depression. Personality traits are important factors that are associated with disorders [21]. For example, prior research has found that conscientiousness and neuroticism are related to insomnia [22,23]. Personality is also the key factor for dropout and treatment resistance in cognitive behavioral therapy for insomnia [24].

Social Media and Insomniac Patterns

A few studies have been conducted to predict insomnia by analyzing the content of social media. Michael et al [10] described a method that uses Twitter for public health research. They exploited the Ailment Topic Aspect Model to create structured disease information from tweets that they used for public health metrics. These authors [25] also reported that early detection of disease outbreaks, medication safety, health behaviors, and individual well-being can be investigated through social media data, and applied traditional natural language processing tools to analyze social media content. Rice et al [26] found that young people may be at risk of negative consequences from new technologies and online media. Therefore, social media platforms are important for measuring young people's mental health states. Andrew et al [27] conducted a study to identify common mental health topics from popular social media platforms and identified common mental health topics such as anxiety, depression, and sleep problems. Jamison-Powell et al [9] completed a study on discussions of insomnia on Twitter. Through an analysis of 18,901 tweets, they discovered that when the word "insomnia" appears in users' tweets, they are likely to convey strong negative health information. These authors mainly conducted their analysis over 2 different themes: coping with insomnia and describing experiences with insomnia. For the first theme, users share symptoms and coping strategies on Twitter, while for the second theme, users share frustration. However, the authors did not build a prediction model and did not explore the association between users' personality and insomnia. McIver et al [28] examined how 2 groups of Twitter users—sleep and nonsleep groups—were active on social media. They found that the nonsleep group showed negative sentiment over social media. Suarez et al [29] conducted a study on the real-time streaming of Spanish tweets for insomnia prediction by analyzing 54,432 tweets and built a classifier whenever insomnia phrases appeared. They used term frequency-inverse document frequency to find features of n-grams and then applied different classifiers including support vector machines and k-nearest neighbors.

It is clear that prior studies suffer from the following deficits in research: prior studies largely built classifiers based on explicit insomnia phrases (eg, "insomnia," "sleepless") rather than the linguistic cues and relationships between the words,

the authors did not build any novel machine learning models, and, to the best of our knowledge, no study predicted insomnia based on users' personalities from social media interactions. The aim of this paper was thus to address these research gaps.

Research Objectives

The main objective of this study is to predict insomnia by analyzing users' psycholinguistic characteristics (eg, word usage patterns) and Big 5 personality traits as derived from their tweets. In this study, we built a rigorous classification model to predict users' insomnia from their interactions on Twitter. We collected a total of 4,198,683 tweets from 1574 users and classified our training data set into 2 classes: individuals who suffer from insomnia and individuals who do not suffer from insomnia. We then conducted psycholinguistic-based analyses of users' tweets by using 2 popular tools: Empath [30] and Linguistic Inquiry and Word Count (LIWC) [31]. We also analyzed users' tweets by using the bidirectional encoder representations from transformers (BERT) [32] word embedding technique to investigate semantic relationships between words. We carefully conducted both psycholinguistic- and word embedding–based analyses to find insights into users' tweets. We then integrated users' personalities with the psycholinguistic models. Finally, we built double-weighted, ensemble-based classification models by combining psycholinguistic-, word embedding–, and personality-based models.

In summary, our study provides the following contributions: a large insomnia data set consisting of 1574 users from different geographical locations; novel insomnia classification models comprising psycholinguistic, word embedding, and Big 5 personality attributes as derived from users' tweets; and a novel and rigorous double-weighted, ensemble-based classification model to predict insomnia.

Methods

In this paper, we performed the 5 following steps to predict insomnia by analyzing users' word use patterns in tweets: (1) user selection—first, we found a total of 1574 users by using Twitter's advanced search technique [33] and divided the users into 2 groups, users with sleep disorders and users with no sleep disorder; (2) linguistic analysis—we performed 2 different types of linguistic analyses, psycholinguistic-based analysis and word embedding–based analysis; (3) correlation analysis—we found correlations between users' psycholinguistic patterns and sleep disorders by using Fisher's latent Dirichlet allocation (LDA) [34]; (4) model building—we built 3 different machine learning models by using psycholinguistic features, word embedding attributes, and the Big 5 personality traits; and (5) ensemble of the models—finally, we built a double-weighted ensemble model by integrating the previous models. These steps are further detailed in the subsequent subsections.

Data Collection

We collected tweets from a total of 1574 users. We searched phrases such as "insomnia" and "sleepless" by using Twitter's advanced search technique. We divided the users into 2 groups: Insomnia Yes and Insomnia No. We then manually annotated the files of the users' tweets with the Insomnia Yes and Insomnia No labels.

We confirmed that randomly selected users were neither organizational nor celebrity profiles. We manually checked whether Insomnia Yes users disclosed their own sleep disorders. We found users' tweets and their locations. The following are a few sample tweets shared by insomniac users on Twitter: "I wish insomnia wasn't a part of my life" and "my insomnia is officially the worst I got into bed at 12, slept for 2 hours, and am still wide awake at 6:30 AM." These tweets indicated that these users suffered from insomnia. If we found several such tweets (28.60 times on average) in the newsfeed or tweets of a user, we labeled that user as Insomnia Yes.

For Insomnia No users, we randomly selected users who did not have these search phrases in their tweets. We also manually checked whether these users shared any sleeping-related issues, and if they did, we discarded them from the list. After labeling Insomnia Yes or No on the users' csv files containing the tweets, we removed text such as "insomnia" and "sleepless" from the users' tweets. By removing these phrases, we created a bias-free data set that contained cues regarding users' insomnia without explicitly mentioning these phrases. Users who used contractions tend to have higher levels of fluency in their pronunciation, while users who do not use contractions are largely nonnative speakers. People may contract many words every day but mainly focus on the word "not" during contraction [35]. For the above reasons, we kept contracted words in our tweets to indicate a difference between the patterns of writing between native and nonnative speakers. Later, we conducted our full analysis again.

It is important to note that we only focused on users who express their insomnia on Twitter. If they are actually suffering from insomnia, then their word usage pattern could be insightful. We collected users' tweets, and multiple referees labeled these users after investigating their writing, their geographic locations, and the irregular timestamps of the tweets. In this way, we could be confident that a given user was suffering from insomnia. We emphasized a users' concern from their tweet content to make decisions about their behavior. It is likely that when users encounter a grave issue, they frequently express their problem to others. Thus, we labeled a user as Insomnia Yes if we found a high frequency (average 28.60) of mentions of the problem. We also found the maximum, minimum, and SD of users' number of posted tweets regarding the aforementioned problem to be 363, 8, and 24.57, respectively. These numbers revealed that some users suffering from extreme insomnia-related problems continue to tweet at nighttime about their sleeplessness.

There might be another group of users who have insomnia but do not disclose the problem in their tweets. We did not consider those users in our study. We also found a large number of users who were noninsomniacs. We again manually checked the users' data and tweet times based on geographic location to confirm that these users were noninsomniacs. If we had users who were noninsomniacs but tweeted in different time patterns to regular users, then we discarded their tweets from our study. As the number of insomniac users is lower in comparison to

noninsomniac users, we took less data of noninsomniac users to make a balanced data set.

It is important to note that there might have been users who had insomnia but do not disclose it in their tweets. However, we argue that they are not in large numbers, as previous studies show that in social media people reveal their actual behavior without idealization [36] and share private traits [37].

Next, we identified the genders of Twitter users based on various attributes to investigate correlations between users' gender and insomniac behavior. First, we manually checked profile pictures and biographies to identify gender, but many users do not share their photos in their profiles. Thus, we observed their writing and the replies of other users to determine their gender from third person pronouns (ie, "he," "she," "him," and "her"). If we could not identify the gender from Twitter, we then checked the other social network accounts of that particular user by conducting a manual search based on their names and usernames. For instance, we could identify the genders of Twitter users based on their pictures on their Instagram or Facebook profiles. If all the above methods failed to ascertain the gender, we discarded the user from our list.

We searched for users from those countries that have English as their first language and were either native or nonnative speakers. We collected users' tweets from a total of 6 different countries: Australia, Canada, Ireland, New Zealand, the USA, and the UK. Table 1 shows the number of users from both groups (Insomnia Yes and Insomnia No) collected from different geographical locations.

Table 2 displays the statistics describing the tweets of our data set.

Table 1. Twitter users' statistics by location.

Countries	Total, n (N=1574)	Insomnia Yes users, n (n=820)	Insomnia No users, n (n=754)
UK	212	108	104
Australia	62	31	31
Canada	334	181	153
USA	919	473	446
New Zealand	32	18	14
Ireland	15	9	6

Table 2. Tweet statistics.

Statistic	Insomnia Yes	Insomnia No
Tweets, n	1,998,683	1,810,567
Maximum number of tweets of a user, n	3247	3250
Minimum number of tweets of a user, n	26	26
Average number of tweets of a user, mean (SD)	2437.42 (1035.42)	2401.28 (1156.65)
Maximum word count of a user, n	67,427	65,660
Minimum word count of a user, n	195	191

For raw data preprocessing, we discarded the username and mentions. We kept the retweets of users because the act of retweeting could indicate the personality traits of users. We removed hashtags (eg, "#insomnia,") and converted them into text (eg, "insomnia"). We also removed URLs and http links because these text data cannot be analyzed by lexical methods. These texts also do not produce any sensible numeric vector for the word embedding method. We did not remove stop words because some word embedding techniques, like BERT, show that prepositions facilitate a better understanding of the context. We removed emojis by using the Python demoji package [38]. For example, we replaced the fire symbol emoji with the word "fire." We used the LIWC2015 dictionary for removing emoticons according to the suggestion of Seabrook et al [39].

Model Building

We built the classification models from 3 different techniques: (1) a psycholinguistic-based model (ie, LIWC and Empath), (2) a word embedding–based model (ie, BERT), and (3) a Big 5 personality–based model. First, we describe the performance of predicting Insomnia Yes and Insomnia No for each of the independent models. We then describe the process of building our novel ensemble model from these 3 independent models. Since we used word-embedding techniques for deep learning–based models, we did not need a feature selection method for this classification category. We carefully selected important features for the other 2 approaches because irrelevant features can weaken the accuracy of a model [40].

Feature Selection

In our data set, our independent variables consisted of analysis scores from users' tweets as generated by psycholinguistic tools (ie, LIWC and Empath), word-embedding methods (ie, BERT), and the Big 5 personality traits. Our dependent variable was Insomnia Yes and Insomnia No. As our independent variables were continuous while the dependent variable was categorical, we applied Fisher' linear discriminant analysis [34] for our feature selection method.

As mentioned earlier, we used 2 different psycholinguistic techniques, LIWC and Empath, to analyze users' tweets. Initially, we used the LIWC-based approach using LIWC2015, which classifies approximately 90 different features from texts into 7 different categories, where each category contains hundreds of words [41]. These categories include summary language variables (analytical thinking, clout, authenticity, and emotional tone), general descriptor categories (words per sentence, percent of target words captured by the dictionary, etc), standard linguistic dimensions (articles, auxiliary verbs, etc), word categories tapping psychological constructs (affect, cognition, etc), personal concern categories (work, home, leisure activities, etc), informal language markers (assents, fillers, swear words, netspeak), and punctuation categories (periods, commas, etc).

We first considered LIWC scores to be independent variables and Insomnia Yes and Insomnia No as dependent variables. We applied Fisher's linear discriminant analysis by using SPSS statistical software (IBM Corp) to find the relevant features. Table 3 shows the correlation coefficient between users' LIWC scores and Insomnia Yes and Insomnia No according to Fisher's linear discriminant analysis. The predictors with larger (>1.0) scores are better predictors. Therefore, these scores are helpful in deciding which variables have more effect when building the classification models [42].

Table 3. Fisher's correlation coefficient between LIWC[a] categories and insomnia categories.

LIWC category	Insomnia Yes	Insomnia No
i	48.117	47.838
negate	123.587	123.253
swear	74.714	74.521
health	18.731	18.466
drives	–31.599	–31.479
focuspresent	17.453	17.342
SemiC	16.856	16.157
cogproc	–38.140	–38.057
sad	19.972	20.991
affiliation	37.245	37.323
anx	17.692	18.524
death	22.654	23.873
social	83.284	83.246
Analytic	–17.013	–16.901

[a]LIWC: Linguistic Inquiry and Word Count.

Second, we used an Empath-based approach to address the shortcomings of the LIWC-based approach. LIWC can only analyze a total of 6400 dictionary words. A dynamic deep learning–based approach was used to analyze these words with Empath. We analyzed the text with Empath by using the empath-client Python implementation package [43]. Empath draws connotations between words and phrases by using deep learning–based neural embedding across more than 1.8 billion words of modern fiction. Given a small set of seed words that characterize a category, Empath uses its neural embedding to discover new, related terms and then validates the category with a crowd-powered filter. Empath analyzed text across 200 built-in, prevalidated categories we generated from common topics in our data set, like neglect, government, and social media. We analyzed the tweets of each user by using Empath and considered the outcome of the tweets as our independent variables; meanwhile, Insomnia Yes or Insomnia No was taken as the dependent variable. Although LIWC and Empath are highly correlated (r=0.906), we found no correlation between Empath and our insomnia classification categories. LIWC has a total of 93 psycholinguistic categories, whereas Empath has a total of 200 word categories. When these 93 LIWC word categories were divided between 200 Empath word categories, which might not have been evenly distributed across the categories, Empath showed weak correlation coefficients. Therefore, we ultimately did not integrate the Empath word categories with our combined model.

Third, to find correlations of users' sleeping patterns with the Big 5 personality traits, we initially computed users' personality scores. We computed the Big 5 personality traits [20] of the users from their tweets by using the IBM Watson Personality Insight API (application programming interface) [44]. Arnoux et al [8] have shown that the IBM Watson Personality Insight API performs well in comparison to other techniques. Other prior studies [45,46] have also used the API and demonstrated reasonable performance. We found that the majority of the traits had a high score of 1.00 and low score of 0.01. We also observed that insomniac users had average openness, conscientiousness, extraversion, agreeableness, and neuroticism scores of 0.61, 0.29, 0.52, 0.56, and 0.83, respectively.

Building Classifiers

After feature selection of all the techniques was completed, we first built a classification model with LIWC. To build the

classification model, we considered 14 different LIWC categories of words as initial features. Table 3 shows the features for the LIWC-based approach. The features are i, negate, swear, health, drives, focuspresent, SemiC, cogproc, sad, affiliation, anx, death, social, and Analytic. A potential problem can arise when collinearity is present among the features. To remove collinearity among independent LIWC features, we computed the correlation among the features by using the R regression subset selection package "leaps" [47]. We found that the following features were collinear: drives, sad, SemiC, focuspresent, and affiliation. We discarded these collinear features from our feature list. Finally, we conducted classification with the relevant features using a 10-fold cross-validation with 10 iterations. We built our classification model with several classifiers that included Naive Bayes, AdaBoost, random forest, support vector machine, and Gaussian process.

After this, we used a linguistic model that is capable of finding contextual relationships between words in a sentence. To this end, we use the BERT [32] model which is pretrained on a large corpus of sentences. The model learns to produce a powerful internal representation of words as embeddings. We used the sentence-transformers [48] library to generate BERT vectors by using a pretrained model. We grouped 2 sets of preprocessed data sets, one without applying lemmatization and punctuation marks and the other after applying all the preprocessing techniques.

We used the BERT vector as input for our convolutional neural network (CNN)–based deep learning model [49]. As CNN is a nonlinear machine learning model and the BERT embedding vector has a large input feature (768 × 1), we trained our model with this deep learning–based architecture. The CNN model contains 2 hidden layers. In the first hidden layer, we added the leaky ReLU (rectified linear unit) [50] activation function. In the next hidden layer, we added a dropout for regularization of the model with a tanh activation function. Finally, we added a dense layer with a softmax activation function [51] and used the Adam optimizer with binary cross-entropy loss. We split the training and test data sets by 70% and 30%, respectively, and built the classification model using a 10-fold cross-validation with 10 iterations. With this configuration, the accuracy was 67% and 58% for the training and test data sets, respectively.

Next, we built our classification model by using the Big 5 personality traits. Based on the date summarized in Table 4, we selected 3 relevant Big 5 traits: conscientiousness, neuroticism, and agreeableness. We built the classification model with the relevant traits using a 10-fold cross-validation with 10 iterations. We built our classification model with several classifiers that included Naive Bayes, AdaBoost, random forest, support vector machine, and Gaussian process by again splitting the training and test data sets by 70% and 30%, respectively.

Table 4. Fisher's correlation coefficient between personality traits and Insomnia Yes and Insomnia No.

Personality trait	Insomnia Yes (Fisher's score)	Insomnia No (Fisher's score)
Conscientiousness	31.885	29.227
Neuroticism	28.168	17.336
Openness	–19.137	–20.785
Extraversion	–1.93	–1.61
Agreeable	2.175	3.132

Building Weighted Ensemble Classifiers

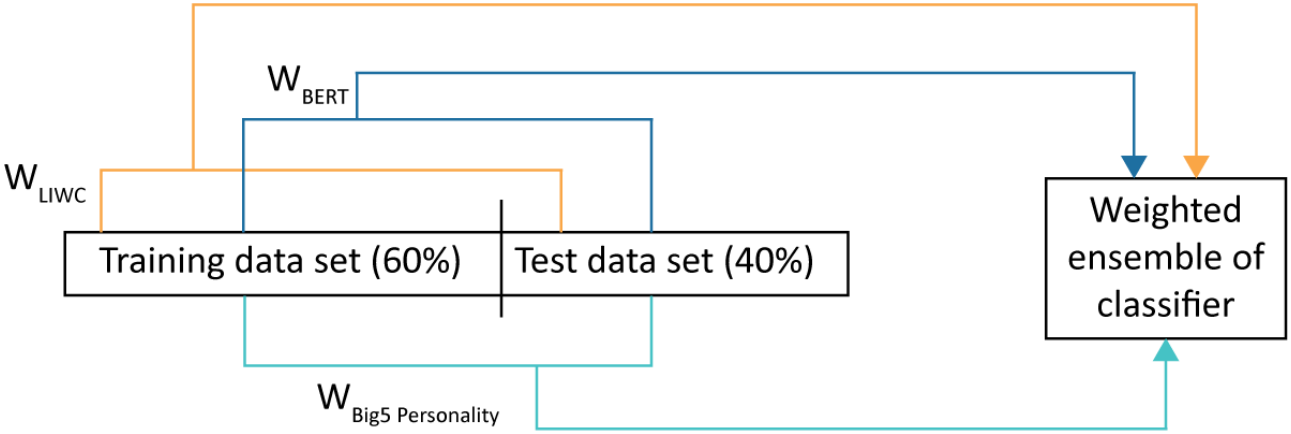
We finally combined the previous classification models to increase the strength of our prediction model. To determine a unified final insomnia label, we combined all the independent models, including, LIWC, BERT, and the Big 5 personality traits. It was necessary to prioritize the approaches based on their performance. We ordered the approaches by assigning weights to each model. We used these weights to build our final ensemble model. To build our ensemble model, we performed the following 2 steps: computing weights of each approach from both the training and test data sets and combining the models with a double-weighted linear ensemble technique.

We then determined the weights of each approach (ie, W_{LIWC}, W_{BERT}, and W_{Big5} personality traits) from both the training and test data sets. We occasionally observed that a model showed better strength with the training data set while the model performed weakly for the test data set. Therefore, we paid close attention to the performance of our models in both the training and test data sets. This way, we could bring greater diversity to the weights to build our ensemble model. To this end, we first ran the classification model over the training data set of 944 Twitter users (944/1574, 60.0% of the total data set). We ran the classification models by using LIWC, BERT, and the Big 5 personality traits. We used both linear (eg, random forest, Naive Bayes) and nonlinear (ie, CNN) models during training. We ranked the strength of these classification models by checking their root mean square error (RMSE) scores. The lower the RMSE score, the higher the weight of a model (linear or nonlinear). To compute the weight, we subtracted the RMSE score from 1 (W=1–RMSE).

We again ran the classification models over the test data set of 630 Twitter users (630/1574, 40.0% of the total data set). We also applied both linear and nonlinear techniques by using LIWC, BERT, and the Big 5 personality trait approaches. We then ranked the weights (1–RMSE for the test data set) of these approaches. Figure 1 shows the detailed process of producing weights from both the training (60%) and test (40%) data sets. In the figure, we indicate weight as W, which indicates that we subtracted RMSE from 1 for the model. In our double-weighted

method, we combined the weights of each type by using convex combination techniques [52] for both the training and test data sets. For the training data set, we obtained weights of 0.52, 0.47, and 0.50 from LIWC, BERT, and the Big 5 traits, respectively. In contrast, for the test data, we achieved weights of 0.38, 0.50, and 0.36 from LIWC, BERT, and the Big 5 traits, respectively. These weights were generated from linear and nonlinear classification models of LIWC, BERT, and the Big 5 personality traits.

Figure 1. Weight computation and building ensemble model for predicting insomnia. BERT: bidirectional encoder representations from transformers; LIWC: Linguistic Inquiry and Word Count.



Finally, we combined the weights that we found from the double-weighted method by using the convex combination method. Equation 1 presents the final insomnia ensemble classification result (I_{Final}) from the previous 3 different models.

where Y_{LIWC}, Y_{BERT}, and Y_{Big5} refer to insomnia prediction results through use of LIWC, BERT, and the Big 5 personality traits, respectively; and W_{LIWC}, W_{BERT}, and W_{Big5} denote the weights generated from LIWC, BERT, and the Big 5 personality traits, respectively, from both the training and test data sets through use of the convex combination technique.

Results

Overview

Our study is the first study to build a novel ensemble learning model to predict insomnia by analyzing a large number of tweets. Prior studies [28,29] investigated the pattern of insomniac behavior by using only a limited number of tweets. Furthermore, the authors explicitly considered phrases such as "insomnia" and "sleepless" in their data sets. In contrast, in our study, we discarded these explicit phrases when predicting users' sleeping issues, which makes our study different and more robust than the previous studies. In this section, we report the performance of our independent and final ensemble-based classifiers, discuss the incorporation of the emotional variability among the insomniac and noninsomniac users from their tweets, and present the distribution of insomniac users and the variability of their Big 5 personality scores. Finally, the correlation between users' genders and their insomniac behavior is discussed.

Performance of Independent and Ensemble Classifiers

First, we investigated the performance of our independent and ensemble-based classifiers. Table 5 shows the performance of the classifiers. We found that the Gaussian process classifier had the best average performance (area under the curve [AUC] 75.3%) in predicting users' insomnia.

Table 5. Strength of the classification model for predicting insomnia by LIWC categories and the Big 5 personality traits.

Classifier, insomnia class	LIWC[a]			Big 5 personality traits		
	TPR[b]	FPR[c]	AUC[d]	TPR	FPR	AUC
Random forest						
Yes	0.716	0.334	0.747	0.686	0.475	0.649
No	0.678	0.270	0.747	0.525	0.314	0.649
Naive Bayes						
Yes	0.740	0.472	0.694	0.746	0.591	0.585
No	0.536	0.260	0.694	0.409	0.254	0.585
SVM [e]						
Yes	0.632	0.266	0.680	0.883	0.674	0.604
No	0.732	0.371	0.680	0.326	0.117	0.604
AdaBoost						
Yes	0.699	0.414	0.694	0.836	0.679	0.599
No	0.579	0.314	0.694	0.321	0.164	0.599
Gaussian process						
Yes	0.747	0.383	0.754	0.765	0.542	0.666
No	0.713	0.376	0.754	0.457	0.234	0.666

[a]LIWC: Linguistic Inquiry and Word Count.

[b]TPR: true-positive rate.

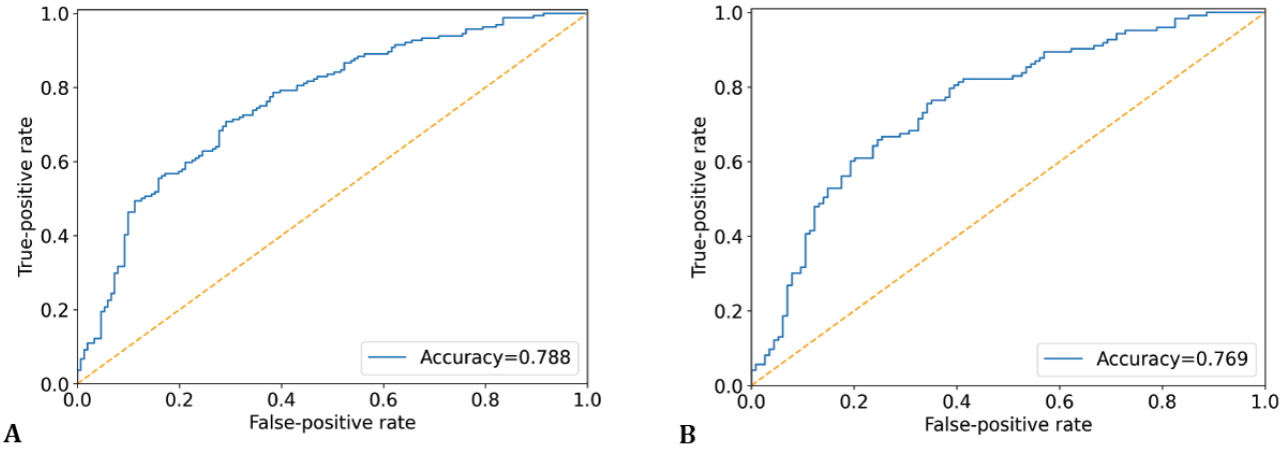
[c]FPR: false-positive rate.

[d]AUC: area under the curve.

[e]SVM: support vector machine.

Our final ensemble classification model for insomnia, I_{Final}, achieved an AUC of 78.8% and 76.91% from the training and test data sets, respectively, outperforming the previous independent models. When observing the performance of the ensemble-based classifiers on the training and test data sets, we found that the performances in the test set were similar to those of the training data set. The receiver operating characteristic curves for the training and test data sets on our ensemble-based classifiers are displayed in Figure 2.

Figure 2. Receiver operating characteristic curves for insomnia classification for the (A) training set and (B) test set.

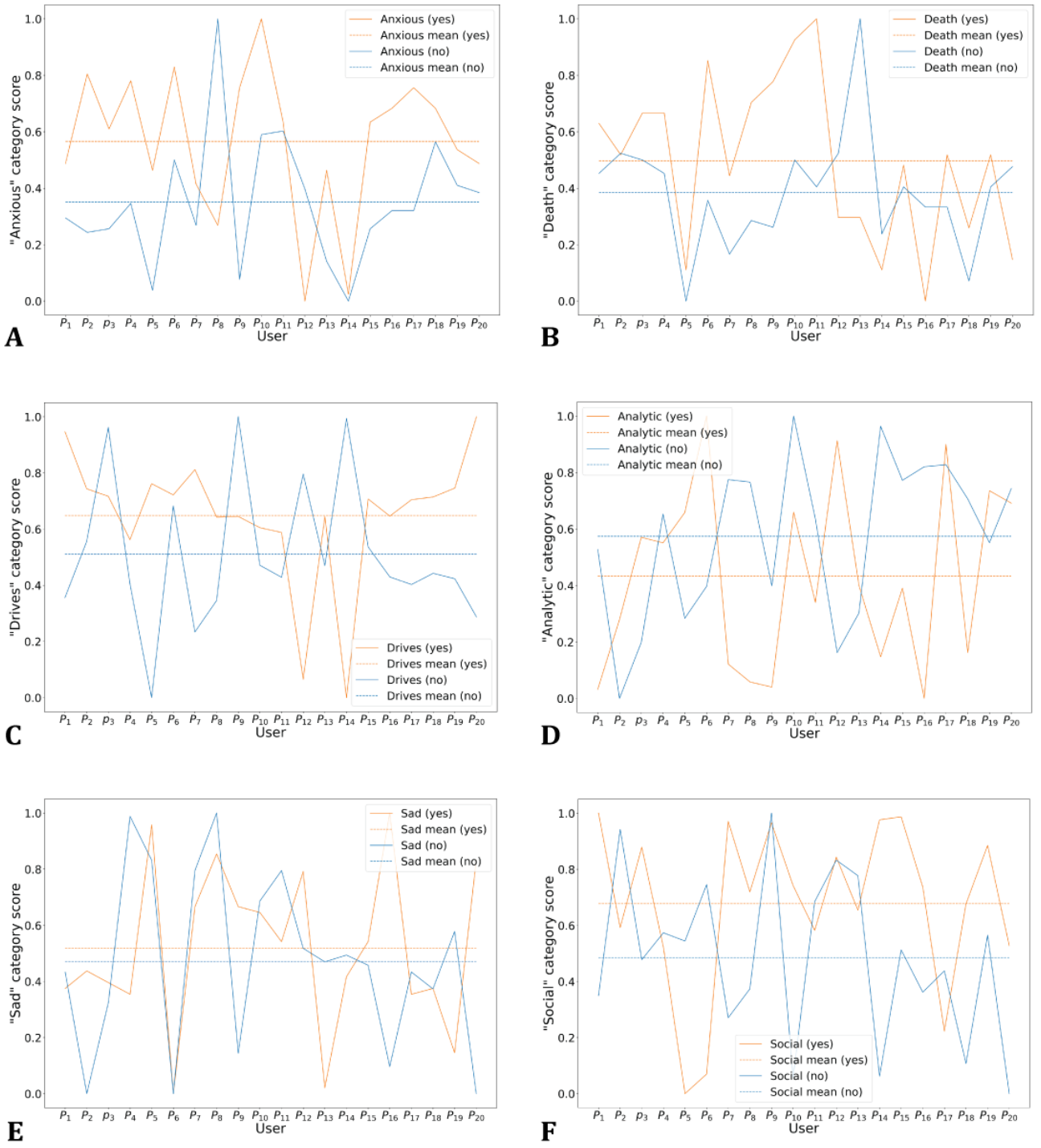


Emotional Variability Among Insomniac and Noninsomniac Users

Our study investigated whether emotional variability exists between insomniac and noninsomniac users based on their tweets. From our observation, we found that users' insomniac behavior and their psycholinguistic categories were correlated. We randomly selected 20 insomniac and 20 noninsomniac users. We extracted users' anxiety-related words, (eg, "worried," "fearful," "nervous," "tense") and converted their scores into a range from 0 to 1 by using the max-min normalization technique. Figure 3 presents the variability in the use of these words

between the insomniac and noninsomniac users. We also observed that on average, insomniac users exploited anxiety-related words 10% more than did the noninsomniac users. Carrera et al [53] showed in their study of 200 college students that a significant association can be found between difficulty sleeping and fear of death. In our study, we also found that insomniac users are likely to post death-related words in their tweets. We observed that insomniac users write more "social words." These users tend to spend more time socializing with others when they face difficulty sleeping. Our study also showed that insomniac users tend to display a lack of analytical thinking.

Figure 3. Usage of words, (A) "Anxious," (B) "Death," (C) "Drives," (D) "Analytic," (E) "Sad," and (F) "Social," related to the LIWC category and their mean scores between insomniac and noninsomniac users. LIWC: Linguistic Inquiry and Word Count.



Visualization of the Correlation Between Big 5 Personality Traits and Insomniac Users

Users' Big 5 personality traits and insomniac behavior showed correlations. Table 6 shows the distributions indicating that agreeableness might have a weak correlation with insomniac behavior. However, conscientiousness had a strong correlation with insomniac behavior (Fisher's score 31.88). The users who had high insomnia were more likely to be more conscientious. In contrast, Wissar et al [54] reported that lower

conscientiousness is associated with greater insomnia severity. We also find that users who had high neuroticism scores were more likely to have severe insomnia. Table 6 shows that insomniac users have moderate scores in agreeableness, are likely to be high in conscientiousness, and are highly neurotic.

Table 6. The distribution of correlated Big 5 personality trait scores among insomniac users.

Traits by range	Percentage, n (%) (N=820)
Agreeableness	
0.0–0.2	77 (9.4)
0.2–0.4	164 (20.0)
0.4–0.6	188 (22.9)
0.6–0.8	214 (26.1)
0.8–1.0	177 (21.6)
Conscientiousness	
0.0–0.2	111 (13.5)
0.2–0.4	213 (26.0)
0.4–0.6	376 (45.9)
0.6–0.8	88 (10.7)
0.8–1.0	32 (3.9)
Neuroticism	
0.0–0.2	9 (1.1)
0.2–0.4	38 (4.6)
0.4–0.6	79 (9.6)
0.6–0.8	125 (15.2)
0.8–1.0	569 (69.4)

Insomnia and Gender Correlation

Seabrook et al [39] showed that gender might have an association with mental health problems, such as depression. Reasoning that insomnia is a mental health problem [55] and motivated by previous studies [39,56], we also investigated the association between users' gender and their insomnia-related problems. We discovered that the number of insomniac and noninsomniac male users was 309 and 363, respectively. We further observed that the number of insomniac and noninsomniac female users was 511 and 391, respectively. From our data set, we conducted chi-square (P<.001) tests based on gender and insomnia where the degrees of freedom ([number of rows–1] × [number of columns–1]) in the contingency table were 1. Our result showed that gender and insomnia were correlated.

After careful observation, we found that among 511 of all female users 6.8% (n=35 users) were suffering from pregnancy- or postpartum-induced insomnia. These users' tweets contain words related to "pregnant," "postpartum," and "cosleep." Later, we applied LDA [57] over the tweets from these users' pregnancy or postpartum times. LDA is a topic modeling technique that automatically organizes a large corpus to discover hidden themes. From users' patterns of tweeting, we could estimate the possible time frame for their pregnancies. Table 7 presents the extracted 5 major topics from their tweets: sleep, night, tired, child, and weird.

Table 7. Distribution of major topics, including (A) sleep, (B) child, (C) night, (D) tired, and (E) weird, extracted from a group of insomniac users' tweets during their pregnancy or postpartum periods.

Topic	Distribution (%)
Topic A	
Sleep	0.085
Problem	0.077
Suffer	0.075
Little	0.045
Anxiety	0.060
Dream	0.057
Escape	0.037
Spiritual	0.036
Random	0.035
Morning	0.030
Topic B	
Birth	0.081
Breast	0.071
Nursing	0.070
Happiness	0.076
Lucky	0.062
Potty	0.079
Scare	0.050
Check	0.048
Patience	0.052
Angel	0.058
Topic C	
Tonight	0.072
Night	0.070
Dinner	0.061
Noise	0.060
Think	0.072
Party	0.048
Event	0.051
Pizza	0.042
Worry	0.050
Carry	0.049
Topic D	
Tired	0.076
Suffer	0.068
Watch	0.061
Stressful	0.060
Minute	0.059
Break	0.070
Hyperemesis	0.043

Topic	Distribution (%)
Pregnancy	0.042
Battle	0.040
Sport	0.039
Topic E	
Weird	0.074
Stare	0.062
Disgust	0.060
Crazy	0.056
Bitch	0.072
Shame	0.050
People	0.070
Employee	0.049
Dislike	0.069
Pregnant	0.068

Discussion

Emotional Linguistic and Insomniac Behavior

As discussed earlier, users' linguistic patterns and Insomnia Yes or Insomnia No indications were strongly correlated. Insomniacs tended to use negative categories of LIWC words (eg, "no," "not," "never"). Some people with strong temperament frequently use swear words (eg, "damn," "piss"). In Bonnet et al's [55] study, insomniacs showed hyperarousal (an abnormal state of responsiveness) due to increased secretion of corticosteroids and adrenaline and an elevated metabolic rate. In another study, Bonnet et al [58] showed that insomniacs experience mood alternation and chronic psychological activation. The following is a sample tweet from an Insomnia Yes user that endorses our finding: "It sucks when you realize that the people closest to you are the most toxic." They also use anxious (eg, worried, fearful, nervous) and sad (eg, crying, grief, sad) categories of LIWC words in their tweets. Freeman et al [59] showed that insomnia has a connection with anxiety, depression, and being worried, and these people are likely to use death-related words (eg, "bury," "coffin," "kill"). Harrison et al [60] reported that insomniacs generally exhibit fear of death and show anxiety about the uncertain (eg, "maybe," "perhaps," "guess"). The following tweet is an example this: "I feel dead and I hate everybody." Our study also adds to these findings by analyzing the correlation between LIWC categories of words found in the users' tweets and their sleeping patterns. Hiller et al [61] showed that cognitive processes (eg, the psychobiological inhibition model) play an important role in understanding and treating insomnia. We obtained a few correlations, such as focuspresent and SemiC, which cannot be intuitively explained.

We did not find a correlation between Empath categories of words and users' sleeping patterns. It is interesting to observe that none of these categories showed any significant correlation between Empath word categories and users' sleeping habits. Fast et al [62] reported that LIWC and Empath word categories are correlated (r=0.906). LIWC has a total of 93 word categories that are distributed across a total of 200 word categories in Empath. When a word category subsumes a greater number of individual words, then the chance of 2 features being correlated, such as LIWC and insomnia class labels, increases. In contrast, when the number of words decreases in a category, such as in Empath, then the possibility of being correlated decreases. In our study, we observed that although LIWC and Empath were correlated, this correlation might not have been sustained in a transitive case. For example, although Empath $\rightarrow$ LIWC and LIWC $\rightarrow$ $Feature_x$ are true, Empath $\rightarrow$ $Feature_x$ does not exist in our data set.

The Big 5 Personality Traits and Insomniac Behavior

We discovered that users' personalities and their sleeping patterns have strong correlations and that users with high neuroticism are more likely to have insomnia. They tend to go through depression, social introversion, repression, and intolerance in terms of somatic health problems. The following is a sample tweet supporting this finding: "Ralph Northam needs to fucking RESIGN already. I have zero tolerance for him." A prior study [63] also reported similar observations related to the correlation between neuroticism and insomnia. Assessing neuroticism may allow for the early detection of catastrophic situations of insomnia. We also observed that users with high conscientiousness scores had strong correlations with insomniac patterns. The following tweet from our data set exemplifies this finding: "I cannot control others actions but I can control mine and my reactions." Larsgaard et al [64] noticed that conscientiousness in people might be correlated with reduced sleeping behavior due to their self-inhibition and the meticulousness in their daily activities.

In our study, we also found a correlation, albeit a weak one, between users' agreeableness and insomnia. The following tweet from an Insomnia Yes user is an example of this: "understand people who talk crap about others thru social media. Listen Linda, why aren't you handling it like an adult?" Dekker et al [65] found that neuroticism, agreeableness, and openness are

directly associated with the insomnia severity index. In our study, we discovered that openness is negatively correlated with users' sleeping patterns. Tsaousis et al [66] also demonstrated that openness and insomnia are negatively correlated with each other. However, several studies [67-70] have demonstrated that openness is unrelated to sleep quality.

In our study, there was negative correlation between extraversion and insomnia. Gray et al [69] found there to be no correlation between users' extraversion and insomniac patterns. Our observations of the relationship between the Big 5 traits and insomnia largely overlap with prior findings [63-65,67-70], which suggest that investigating users' insomnia patterns through social media can leverage our manual effort without asking an individual directly about their sleeping issue. Furthermore, understanding and inspecting one's personality traits may provide clues to underlying causes of vulnerability to developing insomnia.

Pregnancy and Insomniac Behavior

According to our finding, users share their depression during their pregnancies and share their experiences during their postpartum periods. We observed that the major supporting topics were night, suffer, problem, and birth, among others. We also found a few topics that described parents' postpartum-related anxiety. Important topics that we found from users' tweets were birth, breast, tired, suffer, and random, among others. Previous studies also support our findings [71,72].

Conclusions

In this study, we investigated the associations between the psycholinguistic and personality traits of users' with insomnia. We captured both the word usage and semantics of users' tweets through the LIWC and BERT linguistic models and developed 2 models to predict insomnia from LIWC features and BERT word embedding. We then built our third machine learning model that used derived personality traits to predict insomnia from tweets. Finally, we built a rigorous ensemble model by combining the 3 separate models. Our ensemble classifier showed strong prediction potential (AUC 78.8%). The classifier was built by using a novel, double-weighted ensemble technique that outperformed the independent classifiers. We plan to improve our classifier by integrating more data from social networks, such as friend lists, time of tweets, gender, workplace, time spent on activities, etc. We also plan to analyze tweets in different languages.

Conflicts of Interest

None declared.

References

1. Ancoli-Israel S, Krystal A, Pagel J, Neubauer D. The diagnosis and management of insomnia. The Medical Roundtable General Medicine Edition 2012:66-74 [FREE Full text]
2. Roth T. Insomnia: definition, prevalence, etiology, and consequences. J Clin Sleep Med 2007 Aug 15;3(5 Suppl):S7-10 [FREE Full text] [Medline: 17824495]
3. Brower KJ. Alcohol's effects on sleep in alcoholics. Alcohol Res Health 2001;25(2):110-125 [FREE Full text] [Medline: 11584550]
4. Anderson B, Speed E. Social media and health: implications for primary health care providers. University of Essex Research Repository.: Centre for Research in Economic Sociology and Innovation (CRESI); 2010. URL: http://repository.essex.ac.uk/3453 [accessed 2020-10-13]
5. Bennett E. Social media and hospitals: from trendy to essential. In: Futurescan 2001: Healthcare Trends and Implications 2011-2016. Chicago, IL, USA: Society for Healthcare Strategy and Market Development; 2011:43-48.
6. Anderson-Bill ES, Winett RA, Wojcik JR. Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation. J Med Internet Res 2011 Mar 17;13(1):e28 [FREE Full text] [doi: 10.2196/jmir.1551] [Medline: 21441100]
7. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. J Med Internet Res 2007 May 14;9(2):e12 [FREE Full text] [doi: 10.2196/jmir.9.2.e12] [Medline: 17513283]
8. Pierre-Hadrien A, Anbang X, Neil B, Jalal M, Rama A, Vibha S. 25 tweets to know you: a new model to predict personality with social media. 2017 Presented at: 11Th International AAAI Conference on Web and Social Media (ICWSM-17); May 15-17, 2017; Montreal, QC, Canada p. 5513 URL: https://www.icwsm.org/2017
9. Sue J, Conor L, Laura D, Andrew G, Shaun L. "I can't get no sleep": discussing #insomnia on Twitter. : ACM; 2012 Presented at: The SIGCHI Conference on Human Factors in Computing Systems; May 5-12, 2012; New York, NY, USA. [doi: 10.1145/2207676.2208612]
10. Paul MJ, Dredze M. You are what you tweet: Analyzing Twitter for public health. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 5, 2011; Barcelona, Catalonia, Spain p. 265-272.
11. Sullivan SJ, Schneiders AG, Cheang C, Kitto E, Lee H, Redhead J, et al. 'What's happening?' A content analysis of concussion-related traffic on Twitter. Br J Sports Med 2012 Mar;46(4):258-263. [doi: 10.1136/bjsm.2010.080341] [Medline: 21406451]
12. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]

13. van de Laar M, Verbeek I, Pevernagie D, Aldenkamp A, Overeem S. The role of personality traits in insomnia. Sleep Med Rev 2010 Feb;14(1):61-68. [doi: 10.1016/j.smrv.2009.07.007] [Medline: 19897388]
14. Lundh L, Broman J, Hetta J. Personality traits in patients with persistent insomnia. Personality and Individual Differences 1995 Mar;18(3):393-403. [doi: 10.1016/0191-8869(94)00125-c]
15. Coursey RD. Personality measures and evoked responses in chronic insomniacs. J Abnorm Psychol 1975 Jun;84(3):239-249. [doi: 10.1037/h0076716] [Medline: 1133252]
16. Pruthi S. Insomnia. Mayo Foundation for Medical Education and Research. 2016 Oct. URL: https://www.mayoclinic.org/diseases-conditions/insomnia/symptoms-causes/syc-20355167 [accessed 2020-10-15]
17. Arroll B, Fernando A, Falloon K, Goodyear-Smith F, Samaranayake C, Warman G. Prevalence of causes of insomnia in primary care: a cross-sectional study. Br J Gen Pract 2012 Feb 01;62(595):e99-e103. [doi: 10.3399/bjgp12x625157]
18. Morin CM. Insomnia: Psychological Assessment and Management. New York, NY, USA: Guilford Press; 1993.
19. Mukta MSH, Ali ME, Mahmud J. Identifying and validating personality traits-based homophilies for an egocentric network. Social Network Analysis and Mining 2016 Sep 6;6(1):1-16 [FREE Full text] [doi: 10.1007/s13278-016-0383-4]
20. Norman WT. Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. J Abnorm Soc Psychol 1963 Jun;66(6):574-583. [doi: 10.1037/h0040291] [Medline: 13938947]
21. Spielman AJ, Caruso LS, Glovinsky PB. A behavioral perspective on insomnia treatment. Psychiatric Clinics of North America 1987 Dec;10(4):541-553. [doi: 10.1016/s0193-953x(18)30532-x]
22. Akram U, Gardani M, Akram A, Allen S. Anxiety and depression mediate the relationship between insomnia symptoms and the personality traits of conscientiousness and emotional stability. Heliyon 2019 Jun;5(6):e01939 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e01939] [Medline: 31245649]
23. Emert SE, Tutek J, Lichstein KL. Associations between sleep disturbances, personality, and trait emotional intelligence. Personality and Individual Differences 2017 Mar;107:195-200. [doi: 10.1016/j.paid.2016.11.050]
24. Johann AF, Riemann D, Spiegelhalder K. Does perfectionism increase the risk for dropout from cognitive behavioral therapy for insomnia? J Clin Sleep Med 2018 Mar 15;14(3):487-488 [FREE Full text] [doi: 10.5664/jcsm.7012] [Medline: 29510798]
25. Michael JP, Abeed S, John SB, Azadeh N, Matthew S, Karen LS, et al. Social media mining for public health monitoring and surveillance. Biocomputing 2016:468-479. [doi: 10.1142/9789814749411_0043]
26. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
27. Andrew TG, Golab L. Social media mining to understand public mental health. 2017 Presented at: VLDB Workshop on Data Management and Analytics for Medicine and Healthcare; Sept 1, 2017; Munich, Germany URL: https://link.springer.com/chapter/10.1007%2F978-3-319-67186-4_6 [doi: 10.1007/978-3-319-67186-4_6]
28. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing sleep issues using Twitter. J Med Internet Res 2015 Jun 08;17(6):e140 [FREE Full text] [doi: 10.2196/jmir.4476] [Medline: 26054530]
29. Suarez D, Araque O, Iglesias CA. How well do Spaniards sleep? Analysis of sleep disorders based on twitter mining. 2018 Presented at: Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS); Oct 15, 2018; Valencia, Spain p. 11-18. [doi: 10.1109/snams.2018.8554488]
30. Fast E. Ejhfast/empath-client: analyze text with empath. Empath: Understanding Topic Signals in Large-Scale Text.: GitHub; 2016. URL: https://github.com/Ejhfast/empath-client [accessed 2020-11-20]
31. Pennebaker JW, Francis ME, Booth RJ. Linguistic Inquiry and Word Count: LIWC. Mahwah, NJ, USA: Lawrence Erlbaum Associates; 2001.
32. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: North American Chapter of the Association for Computational Linguistics; June 2-7, 2019; Minneapolis, MN, USA URL: https://aclanthology.org/N19-1423
33. Twitter Advanced Search. Twitter. URL: https://twitter.com/search-advanced [accessed 2020-07-20]
34. Mika S, Ratsch G, Weston J, Scholkopf B, Mullers KR. Fisher discriminant analysis with kernels. 1999 Presented at: The 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468); Aug 25, 1999; Madison, WI, USA p. 41-48. [doi: 10.1109/nnsp.1999.788121]
35. Rules for Compositors and Readers at the University Press, Oxford. Oxford, UK: Oxford University Press; Apr 1928:533-533.
36. Back MD, Stopfer JM, Vazire S, Gaddis S, Schmukle SC, Egloff B, et al. Facebook profiles reflect actual personality, not self-idealization. Psychol Sci 2010 Mar;21(3):372-374. [doi: 10.1177/0956797609360756] [Medline: 20424071]
37. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
38. demoji. Python Package Index. URL: https://pypi.org/project/demoji [accessed 2020-10-27]
39. Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 May 08;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]

40. Brownlee J. Feature selection. In: Data Preparation for Machine Learning, Data Cleaning, Feature Selection, and Data Transforms in Python. San Francisco, CA, USA: Machine Learning Mastery; 2019:110-211.
41. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. UT Faculty/Researcher Works. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2020-06-25]
42. Interpreting results of discriminant analysis. OriginLab Corporation. 2016. URL: https://www.originlab.com/doc/Origin-Help/DiscAnalysis-Result [accessed 2020-10-13]
43. empath. Python Package Index. URL: https://pypi.org/project/empath [accessed 2020-12-27]
44. IBM Watson Personality Insights service. IBM Cloud. URL: https://cloud.ibm.com/docs/personality-insights [accessed 2021-03-27]
45. Liu Z, Wang Y, Mahmud J, Akkiraju R, Schoudt J, Xu A, et al. To buy or not to buy? Understanding the role of personality traits in predicting consumer behaviors. : Springer; 2016 Presented at: International 8th Conference on Social Informatics; Nov 11, 2016; Bellevue, WA, USA p. 337-346. [doi: 10.1007/978-3-319-47874-6_24]
46. Mukta MSH, Khan EM, Ali ME, Mahmud J. Predicting movie genre preferences from personality and values of social media users. 2017 Presented at: The Eleventh International AAAI Conference on Web and Social Media; May 15, 2017; Montréal, QC, Canada p. 624-627 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14910
47. CRAN Package leaps. Comprehensive R Archive Network (CRAN) by R Foundation. URL: https://cran.r-project.org/web/packages/leaps/index.html [accessed 2020-12-21]
48. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. : Association for Computational Linguistics; 2019 Presented at: The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November 3-7, 2019; Hong Kong p. 3980-3990. [doi: 10.18653/v1/d19-1410]
49. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. : Association for Computational Linguistics; 2014 Presented at: The 52nd Annual Meeting of the Association for Computational Linguistics; June 22-27, 2014; Baltimore, MD, USA p. 655-665 URL: https://aclanthology.org/P14-1062 [doi: 10.3115/v1/p14-1062]
50. Zhang X, Zou Y, Shi W. Dilated convolution neural network with leakyrelu for environmental sound classification. 2017 Presented at: 22nd International Conference on Digital Signal Processing (DSP); August 23-25, 2017; London, UK. [doi: 10.1109/icdsp.2017.8096153]
51. Li X, Chang D, Tian T, Cao J. Large-margin regularized softmax cross-entropy loss. IEEE Access 2019;7:19572-19578. [doi: 10.1109/access.2019.2897692]
52. Chanaim S, Sriboonchitta S, Rungruang C. A convex combination method for linear regression with interval data. : Springer; 2016 Presented at: International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making; Nov 30-Dec 2, 2016; Da Nang, Vietnam p. 469-480 URL: http://www.jaist.ac.jp/IUKM/IUKM2016 [doi: 10.1007/978-3-319-49046-5_40]
53. Carrera RN, Elenewski JJ. Implosive therapy as a treatment for insomnia. J. Clin. Psychol 1980 Jul;36(3):729-734. [doi: 10.1002/1097-4679(198007)36:3<729::aid-jclp2270360321>3.0.co;2-f]
54. Rojo-Wissar, Wanigatunga AA, Simonsick EM, Terracciano A, Wu MN, Zipunnikov V, et al. Personality and insomnia symptoms in older adults: the Baltimore Longitudinal Study of Aging. Sleep 2021 Sep 13;44(9):e. [doi: 10.1093/sleep/zsab082] [Medline: 33791794]
55. Bonnet M, Arand D. Hyperarousal and insomnia. Sleep Medicine Reviews 1997 Dec;1(2):97-108. [doi: 10.1016/s1087-0792(97)90012-5]
56. 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]
57. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. The Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]
58. Bonnet MH, Arand DL. Caffeine use as a model of acute and chronic insomnia. Sleep 1992;15(6):526-536. [doi: 10.1093/sleep/15.6.526]
59. Freeman D, Stahl D, McManus S, Meltzer H, Brugha T, Wiles N, et al. Insomnia, worry, anxiety and depression as predictors of the occurrence and persistence of paranoid thinking. Soc Psychiatry Psychiatr Epidemiol 2012 Aug;47(8):1195-1203. [doi: 10.1007/s00127-011-0433-1] [Medline: 21928153]
60. Harrison GE, Van Haneghan JP. The gifted and the shadow of the night: Dabrowski's overexcitabilities and their correlation to insomnia, death anxiety, and fear of the unknown. Journal for the Education of the Gifted 2011 Oct 01;34(4):669-697. [doi: 10.1177/016235321103400407]
61. Hiller RM, Johnston A, Dohnt H, Lovato N, Gradisar M. Assessing cognitive processes related to insomnia: A review and measurement guide for Harvey's cognitive model for the maintenance of insomnia. Sleep Med Rev 2015 Oct;23:46-53. [doi: 10.1016/j.smrv.2014.11.006] [Medline: 25645129]
62. Fast E, Chen B, Bernstein MS. Empath: Understanding topic signals in large-scale text. New York, NY, United States: Association for Computing Machinery; 2016 Presented at: The 2016 CHI Conference on Human Factors in Computing

Systems; May 7-12, 2016; San Jose, CA, USA p. 4647-4657 URL: https://hci.stanford.edu/publications/2016/ethan/empath-chi-2016.pdf [doi: 10.1145/2858036.2858535]
63. Taylor DJ, Lichstein KL, Durrence HH. Insomnia as a health risk factor. Behavioral Sleep Medicine 2003 Nov;1(4):227-247. [doi: 10.1207/s15402010bsm0104_5]
64. Larsgård B, Saksvik-Lehouillier I. The predictive power of personality traits on insomnia symptoms: A longitudinal study of shift workers. Personality and Individual Differences 2017 Sep;115:35-42. [doi: 10.1016/j.paid.2016.08.017]
65. Dekker K, Blanken T, Van Someren E. Insomnia and personality-a network approach. Brain Sci 2017 Mar 02;7(3):1-12 [FREE Full text] [doi: 10.3390/brainsci7030028] [Medline: 28257084]
66. Tsaousis I. Circadian preferences and personality traits: A meta-analysis. Eur J Pers 2020 Dec 02;24(4):356-373. [doi: 10.1002/per.754]
67. Stephan Y, Sutin AR, Bayard S, Križan Z, Terracciano A. Personality and sleep quality: evidence from four prospective studies. Health Psychol 2018 Mar;37(3):271-281 [FREE Full text] [doi: 10.1037/hea0000577] [Medline: 29172602]
68. DeYoung CG, Hirsh JB, Shane MS, Papademetris X, Rajeevan N, Gray JR. Testing predictions from personality neuroscience. Brain structure and the big five. Psychol Sci 2010 Jun;21(6):820-828 [FREE Full text] [doi: 10.1177/0956797610370159] [Medline: 20435951]
69. Gray EK, Watson D. General and specific traits of personality and their relation to sleep and academic performance. J Pers 2002 Apr;70(2):177-206. [doi: 10.1111/1467-6494.05002] [Medline: 11908845]
70. Jackson LA, Gerard DA. Diurnal Types, the "Big Five" Personality Factors, and Other Personal Characteristics. Journal of Social Behavior and Personality 1996;11(2):273.
71. De Chaudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 Presented at: The SIGCHI Conference on Human Factors in Computing Systems; Apr 27-May 3, 2013; Paris, France. [doi: 10.1145/2470654.2466447]
72. De Choudhury M, Counts S, Horvitz EJ, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. : ACM; 2014 Presented at: The 17th ACM conference on Computer supported cooperative work & social computing; Feb 15-19, 2014; Baltimore, MD, USA p. 626-638 URL: https://dl.acm.org/doi/10.1145/2531602.2531675 [doi: 10.1145/2531602.2531675]

Abbreviations

API: application programming interface
AUC: area under the curve
BERT: bidirectional encoder representations from transformers
CNN: convolutional neural network
LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
ReLU: rectified linear unit
RMSE: root mean square error

Edited by R Kukafka, G Eysenbach; submitted 30.01.21; peer-reviewed by M Antoniou, PH Makkonen; comments to author 17.03.21; revised version received 12.05.21; accepted 05.10.21; published 09.12.21.

Please cite as:
Sakib AS, Mukta MSH, Huda FR, Islam AKMN, Islam T, Ali ME
Identifying Insomnia From Social Media Posts: Psycholinguistic Analyses of User Tweets
J Med Internet Res 2021;23(12):e27613
URL: https://www.jmir.org/2021/12/e27613
doi:10.2196/27613
PMID:34889758

Original Paper

Factors Influencing Willingness to Share Health Misinformation Videos on the Internet: Web-Based Survey

Alla Keselman[1], PhD; Catherine Arnott Smith[2], PhD; Gondy Leroy[3], PhD; David R Kaufman[4], PhD

[1]Office of Engagement and Training, National Library of Medicine, Bethesda, MD, United States

[2]Information School, University of Wisconsin Madison, Madison, WI, United States

[3]Management Information Systems, University of Arizona, Tucson, AZ, United States

[4]Medical Informatics Program, SUNY Downstate Health Sciences University, Brooklyn, NY, United States

Corresponding Author:
Alla Keselman, PhD
Office of Engagement and Training
National Library of Medicine
8600 Rockville Pike
Bethesda, MD, 20894
United States
Phone: 1 301 827 5671
Email: keselmana@nih.gov

Abstract

Background: The rapidly evolving digital environment of the social media era has increased the reach of both quality health information and misinformation. Platforms such as YouTube enable easy sharing of attractive, if not always evidence-based, videos with large personal networks and the public. Although much research has focused on characterizing health misinformation on the internet, it has not sufficiently focused on describing and measuring individuals' information competencies that build resilience.

Objective: This study aims to assess individuals' willingness to share a non–evidence-based YouTube video about strengthening the immune system; to describe types of evidence that individuals view as supportive of the claim by the video; and to relate information-sharing behavior to several information competencies, namely, information literacy, science literacy, knowledge of the immune system, interpersonal trust, and trust in health authority.

Methods: A web-based survey methodology with 150 individuals across the United States was used. Participants were asked to watch a YouTube excerpt from a morning TV show featuring a wellness pharmacy representative promoting an immunity-boosting dietary supplement produced by his company; answer questions about the video and report whether they would share it with a cousin who was frequently sick; and complete instruments pertaining to the information competencies outlined in the objectives.

Results: Most participants (105/150, 70%) said that they would share the video with their cousins. Their confidence in the supplement would be further boosted by a friend's recommendations, positive reviews on a crowdsourcing website, and statements of uncited effectiveness studies on the producer's website. Although all information literacy competencies analyzed in this study had a statistically significant relationship with the outcome, each competency was also highly correlated with the others. Information literacy and interpersonal trust independently predicted the largest amount of variance in the intention to share the video (17% and 16%, respectively). Interpersonal trust was negatively related to the willingness to share the video. Science literacy explained 7% of the variance.

Conclusions: People are vulnerable to web-based misinformation and are likely to propagate it on the internet. Information literacy and science literacy are associated with less vulnerability to misinformation and a lower propensity to spread it. Of the two, information literacy holds a greater promise as an intervention target. Understanding the role of different kinds of trust in information sharing merits further research.

(J Med Internet Res 2021;23(12):e30323) doi:10.2196/30323

KEYWORDS

misinformation; information literacy; science literacy; webcasts as topic; YouTube

Introduction

Introducing the Concern

When it comes to digital consumer health information, web-based resources can provide both valuable information and misinformation [1]. Concerns regarding the quality of health information and individuals' susceptibility to misinformation are as old as health information websites that target consumers. The changing digital ecosystem of the social media era has increased the potential for and the speed of spread of misinformation, making these concerns more urgent. At present, web-based health information exists in many different formats, often with videos and interactive features. Entire networks, such as YouTube, enable the sharing of videos in an easy manner accessible to a broad public. Social networks allow consumers to spread links to sources that promise quick and easy remedies and promote wellness approaches that are not supported by evidence.

A common theme in non–evidence-based consumer health information is strengthening the immune system and developing natural immunity against infectious diseases. This topic relates to many important public health concerns, from vaccination compliance to individual behavior during epidemics. The COVID-19 pandemic and the need for pandemic-curbing measures underscore the importance of understanding how the public evaluates and shares information about non–evidence-based remedies. We are writing this paper approximately a year after the pandemic reached the United States, and the word *infodemic* is often used to characterize the accompanying epidemic of unreliable information about the virus and the disease [2].

An extensive body of research suggests that individuals often lack the information skills required for navigating effectively through a multitude of health information sources of varying formats and quality [3]. To the best of our knowledge, no studies to date have attempted to characterize the relationship between an individual's tendency to share or recommend a video with non–evidence-based health advice and multiple cognitive and social factors. Understanding how such factors may affect information sharing is essential to curtail the spread of misinformation. This study is concerned with the types of statements that individuals view as convincing support of health claims, as well as the relationship between several factors—information literacy, science literacy, knowledge of the immune system, interpersonal trust, and trust in health authority—and inclination to share a non–evidence-based claim.

A Word on Misinformation

This study focuses on a YouTube video that promotes a dietary supplement, making strong claims about this supplement's ability to *boost* the immune system. Research into how vitamins, minerals, enzymes, and other substances may modulate the immune system is highly complex. However, the claims made by the video are simple and confident. While discussing these claims, we use the term *misinformation*, hence not making any assumptions about the intent of the supplement's producers and promoters.

Misinformation has been defined as *objectively incorrect information that is not supported by scientific evidence and expert opinion* [4]. Although the 2016 US presidential election brought intense attention to the phenomenon, Molina et al [5] analogized it to a much older problem—*yellow journalism*—dating from the Spanish-American War of 1896. It differs from disinformation in that disinformation connotes "an intentional, deliberate, or purposeful effort to mislead, deceive, or confuse."

Wardle and Derakhshan [6] considered information problems for journalists and outlined 7 types of misinformation and disinformation. In this taxonomy, problematic content is contrasted with genuine content:

1. Misleading content (misleading use of information)
2. Satire or parody (without intention to harm)
3. Fabricated content (false content intended to deceive)
4. Imposter content (impersonation of genuine sources)
5. Manipulated content (manipulation of genuine sources to deceive)
6. False context (attached to genuine content)
7. False connection (mismatch of headline to content with intent to deceive).

Clearly, some of these problematic content types are more malicious than others. However, whether intentional or not, misinformation can cause harm. This paper focuses on health information that makes unsubstantiated and exaggerated claims, purposefully or not, with the ultimate objective being to promote and sell a product.

Health Misinformation on the Internet

The prevalence of health misinformation is an ongoing concern. Keselman et al [7] analyzed 24 type 2 diabetes top result pages, obtained from a Google search on *diabetes*, *reversal*, and *natural*. Most of the sites either promised or implied full recovery, most commonly achieved by taking dietary supplements, making claims that opposed the evidence-based perspective of the American Diabetes Association.

Research into behavior on the internet shows that people are susceptible to misinformation and are willing to share it. Web-based health misinformation has been investigated in various places, including web-based health communities [6], social media [8,9], Google Trends [10,11], Amazon's web-based bookstore [12], and smart assistants such as Apple Siri and Amazon Echo [13].

Most research into how people spread misinformation in web-based social networks has been conducted in the domains of vaccines and infectious diseases [14]. For example, Basch et al [15] analyzed the most viewed vaccine-related videos on YouTube. The most popular videos were antivaccine (65.5%). They were commonly posted by laypeople and most often focused on the causal links between vaccines and autism. This and other such studies [16,17] not only illustrate the prevalence of misinformation but also underscore its potential impact. Krishna found that "knowledge-deficient, vaccine-negative individuals" were more active in spreading messages on the internet about vaccines [18]. The messages they spread were aligned with their attitudes.

In a study that focused on social media spread, Chen et al [19] analyzed 2691 tweets on Weibo (a Chinese equivalent of Twitter) related to breast and cervical cancers. They found that about 30% of the tweets contained misinformation, and that treatment-related tweets were more likely to contain misinformation than prevention-related tweets.

Sommariva et al [8] investigated stories circulating in social networking services about the Zika virus and found that half of the *top 10* were either misinformation or rumors. Studies are now beginning to emerge, focusing on the exchange of COVID-19 misinformation. Two-thirds of Korean adults surveyed by Lee et al [20] reported encountering myths and misinformation about COVID-19. Scientific discoveries, especially when they are of focal public interest and rapidly disseminated, beget misinformation; COVID-19 vaccines are being developed so fast that acquiring enough information about them to make decisions is a difficult task. In the United States, in May 2020, 73% of Republicans and 82% of Democrats surveyed called COVID-19 misinformation "a major problem," and almost 70% of those surveyed identified social media as the principal source of such misinformation [9].

Information Competencies

Overview

Researchers have been looking for effective approaches to promote accurate information in impactful ways and for factors that may reduce individual vulnerability. One of the most researched factors is health literacy, sometimes defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions." [21]. Other cognitive factors, while emerging as important theoretical constructs, have been studied less often in an empirical context. Sometimes these constructs are merged with the concept of health literacy (for a review of the conceptual underpinnings and research gaps in health literacy [22,23]). Three of these constructs—information literacy, science literacy, and trust—were considered in this study.

Information Literacy

Pinto et al reviewed the evolution of information literacy as a term and concept [24]. One useful definition is that information literacy is "a set of aptitudes to locate, handle, evaluate and use information efficiently for a wide variety of purposes, generating competent citizens who are better able to perform their activities in the new society" [24].

Information literacy instruction originated in libraries, in the practice known as *bibliographic instruction*. For this reason, both instruction and evaluation techniques were developed for *static, primarily print, sources* [24] and not the information content encountered by today's social media users. For this reason, the assessment of information literacy is typically conducted by academic librarians engaged in the instruction of a specific community of learners—from kindergarten through college—to determine that those learners are meeting the goals of a specific library instruction program. Instruments are developed by librarians and for libraries to fit their institutional missions. Information literacy is an *essential learning outcome* according to the Association of American Colleges and Universities [25].

Information literacy skills apply to any domain of information, including but not limited to health. An information-literate person is one who has "learned how to learn" [26]. That person can find appropriate information resources to meet their specific needs, can search those information resources effectively to find information, and distinguish between different types of resources based on characteristics that include quality markers. Health literacy, in contrast, relates specifically to outcomes affecting the health of the individual, for example, the ability to function in the health care environment (Berkman et al [27] presents a review of health literacy definitions).

Research assessing the effectiveness of information literacy is difficult to find, in part, because there are *very few* measures of the application of real-world learning [28]. However, the importance of information literacy for effectively navigating information and avoiding misinformation is in the very definition of the concept of information literacy.

Science Literacy

Researchers in the field of science education describe scientific literacy as comprising 3 components: (1) knowledge of science, (2) knowledge about science, and (3) attitudes toward science [29].

Knowledge of science refers to understanding specific scientific concepts, for example, being able to explain how antibodies bind to antigens. Research into the impact of knowledge of science on daily life has been very limited, and the few existing studies have produced mixed results. Layton et al [30] studied how people make decisions in situations where science has a potential bearing, for example, the older adults planning their heating budgets. They found that people rarely framed their problems as scientific ones, attending more to social and emotional aspects (eg, the pleasure of a glowing heater). In contrast, Keselman et al [31] found that the depth of biological knowledge contributed to adolescents' ability to reject myths about HIV. Overall, there is no convincing evidence that the knowledge of science is a ubiquitous and powerful aid in real-life situations. In part, the uncertain role of knowledge may be attributed to the depth of knowledge required. More recently, Keselman et al [7] evaluated websites promoting remedies for type 2 diabetes and found that the treatment mechanisms described on these sites referred to concepts well beyond high school biology.

Knowledge about science refers to understanding what constitutes a scientific question, how science may go about deriving the answer, and the checks and balances that exist to minimize bias and reduce errors [29]. Knowledge about science is a potentially promising competency that may be related to evaluating health information on the internet. Individuals with this competency may be skeptical about conspiracy claims and promises of simple solutions for complex problems and attentive to egregious research design flaws. Although science education recognizes the importance of knowledge about science [32,33], empirical investigations are complicated by the absence of operational definitions and assessment tools. Few instruments

isolate this component from general science literacy; the one instrument we found (and used as a basis for this study) targets students in college-level science classes [34].

The third component, attitude toward science, incorporates the view that science is valuable as an advantageous way to derive knowledge about the natural world: understanding the importance of the scientific process and trusting science as an enterprise. Trust in science can undeniably influence individuals' beliefs, consequential to matters of health. Agley and Xiao conducted a cross-sectional survey on the believability of 5 COVID-19 narratives, ranging from *scientific consensus* to *conspiratorial* [35]. Trust in science was the only significant predictor of the acceptance of more scientifically grounded narratives. Jennings et al [36] also found that trust played a role in the willingness to get vaccinated and the rejection of conspiracies.

While discussing science literacy throughout the rest of this paper, we use the term to refer to the *knowledge about science* component, as a generalizable aspect of science literacy that can be addressed through science education. When we write about knowledge of science, we refer to knowledge of the immune system, as the area of science knowledge that is directly applicable to the subject of our study.

In everyday situations that require weighing evidence, science literacy may be overridden by other factors. One factor is sociocultural. The cultural cognition thesis [37] proposes that individual perceptions of societal risks cohere with values characteristic of groups with which they identify. People process information in a manner adjusted to preserve identity-consistent rather than factually correct beliefs [36]. In a large-scale study evaluating views on climate change, Kahan et al [38] found that participants with the most sharply polarized cultural views, not those with high scientific literacy, were most concerned about climate change.

Pennycook et al [39,40] put forth a different kind of argument. They draw on dual-process theory [41] to distinguish System 1 thinking—fast, intuitive, and effortless—from System 2 thinking—analytic and effortful. Pennycook et al [42] argue that lazy System 1 thinking, more than strong partisan belief, predisposes individuals to believe in fake news. People tend to share false health claims from social media partly because they fail to sufficiently consider whether the content is accurate [42]. Moreover, different information presentation formats tend to trigger different systems, with film, compared with graphs and formulas, being more likely to activate analytical reasoning [43].

Trust

In social science literature, the concept of interpersonal trust is a measure of social capital that reflects the goodwill and sense of community that members of a society feel toward one another [44]. Higher general trust is a sign of well-functioning democratic societies and cohesive communities [45]. Multiple studies have found that trust is positively correlated with self-reported health and happiness [45]; trust is viewed as a positive, desirable characteristic of individuals in societies. Generalized trust also correlates with institutional trust in the government and its foundational institutions. Rotter distinguished between *trust* and *gullibility* by defining trust as "believing others in the absence of a clear-cut reasons to disbelieve" [46]. However, to the best of our knowledge, no study has investigated the relationship between interpersonal trust and web-based information behavior. In developing this study, we were interested in understanding the relationship among interpersonal trust, trust in health authorities (eg, leading biomedical research organizations, public health authorities, health care providers) and susceptibility to web-based health misinformation.

Objectives

The study described in this paper has 3 specific objectives:

1. Assess individuals' willingness to share a non–evidence-based YouTube video about strengthening the immune system and characterize types of evidence that positively or negatively affect the opinion of a non–evidence-based remedy.
2. Describe levels of information literacy, science literacy, knowledge of the immune system, and trust among the participants.
3. Characterize the relationship between information literacy, science literacy, knowledge of the immune system, trust, and the likelihood of recommending a YouTube video about a supplement with an immunity-boosting claim.

Methods

Participants

Participants (N=150) were recruited via Regional Medical Libraries (RMLs) of the National Network of the Libraries of Medicine (NNLM). NNLM is a diverse network of 7186 academic health sciences libraries, other special biomedical libraries, public libraries, information centers, and community-based organizations. RMLs coordinate the Network's operations in their regions, supporting and coordinating regional and national programs focused on health information. One coauthor emailed the 8 RML Associate and Executive Directors in late November of 2020 and requested permission to send out the study recruitment announcement to public libraries and community organizations using the listservs in their networks. Listserv members—employees of NNLM member organizations—were asked to promote the survey link among their patrons or clients via means appropriate to their environment. We hoped that this recruitment method would increase the likelihood of the geographic and cultural diversity of the participants. The announcement specified that participants needed to be proficient in English and comfortable using the internet. Professional librarians were excluded from the study.

The initial objective was to recruit 100 participants. Although we did not have previous data that could serve as the basis for power analysis to estimate the needed sample size, past experience suggested that this number would provide a sufficiently large sample size to answer our inferential questions; however, it would also be small enough for a thorough review of open-ended questions that would enrich the study. The first

response was received on December 2. On December 3, the survey was closed because the target number had been reached.

Participants came from 34 US states, including Alaska and Hawaii, and from the District of Columbia. Demographic characteristics of the participants are presented in Table 1.

Upon completing the survey, participants received a US $50 Amazon e-gift card. The study protocol was exempted from Institutional Review Board's review by the National Institutes of Health Office of Human Subjects Research.

Table 1. Demographic characteristics of the participants (N=149)[a].

Demographic characteristics	Participants, n (%)
Age (years)	
18-29	35 (23.5)
30-49	87 (58.4)
50-64	24 (16.1)
65+	3 (2)
Education	
High school or less	12 (8.1)
Some college	28 (18.8)
College graduate	83 (55.7)
Postgraduate degree	26 (17.4)
Gender	
Female	70 (47)
Male	79 (53)
Race and ethnicity	
White	89 (59.7)
Black or African American	21 (14.1)
Hispanic or Latino	17 (11.4)
Asian	5 (3.4)
American Indian or Alaska Native	10 (6.7)
Native Hawaiian or Other Pacific Islander	2 (1.3)
Declined to answer	7 (4.7)
Residential setting	
Urban	102 (68.5)
Rural	11 (7.4)
Suburban	36 (24.2)

[a]One participant did not complete the demographic section.

Instruments

Participants completed a web-based survey comprising the following components (Multimedia Appendix 1).

Web-Based Health Information Evaluation Task

For this task, participants viewed a 5-minute YouTube video, promoting a dietary supplement to strengthen the immune system [47].

Participants received the following prompt:

> *Your cousin had a bad case of the flu last year and did not fully recover for several weeks. In recent years, your cousin has frequently suffered from cold and occasionally flu. A friend told you about how certain supplements can help rev up your immune system. That friend recommends the following video: (link to the video).*

The video is an excerpt from the 2019 TMJ4-TV News episode of *The Morning Blend* TV show, in which 2 anchors talk to a Welltopia Pharmacy pharmacist about a cold and flu-fighting dietary supplement Viracid, which is claimed to "boost your immune system like no other." The episode is sponsored by Welltopia Pharmacy, and the sponsorship is clearly indicated in the video caption. The studio table contains multiple bottles of Viracid. On the close-up, viewers can see the price tag of US $37.50 on a bottle with 60 capsules. The pharmacist recommends taking one capsule of the supplement every hour at the onset of cold and flu symptoms. The pharmacist claims

that the ingredients are "very pure," "in concentrated form," and "the best immune booster" on the market. The video offers a 10% discount when buying the supplement.

After viewing the video, the participants answered several questions. The first asked them how likely they were to recommend the video to their cousin. The 4-point Likert score answers ranged from *Very unlikely* to *Very likely*. Participants were also asked to explain their decision and comment in free text on the convincing and concern-raising aspects of the video. Next, they were presented with several hypothetical pieces of evidence and asked how likely that evidence was to affect their opinion about the supplement. The answer options were *Very likely, negatively*; *Somewhat likely, negatively*; *No effect*; *Somewhat likely, positively*; and *Very likely, positively*. The hypothetical pieces of evidence are as follows:

- A friend taking this supplement for the past 2 years had not had a cold.
- The supplement's rating on a crowdsourced review site is 4.5/5.
- The supplement's rating on a crowdsourced review site is 2.3/5.
- A survey of consumers of the product reported that 85% did not have the flu within a year.
- Positive reviews on the supplement producer's website.
- The explanation of the supplement's biochemical mechanism of action on the company's website.
- Knowledge that the supplement has not been evaluated in a clinical trial.
- The National Institute of Allergy and Infectious Diseases (NIAID) review finding inconclusive evidence about the effectiveness of the supplement.
- NIAID review finding evidence of the effectiveness of the supplement in reducing the frequency and severity of respiratory infections.

Participants had the opportunity to explain their ratings.

Demographic Survey

The demographic survey included questions about participants' gender, age, race and ethnicity, highest level of education achieved, and place of residence.

Information Literacy Survey

The information literacy survey, developed by the authors, comprised 6 multiple-choice questions. The questions were modeled on skill domains and questions presented in 2 existing instruments: *Test of Scientific Literacy Skills* (TOSLS) [34] (refer to the *Science Literacy Survey* section) and *Parenting Plus Skills Index* [48]. The *Parenting Plus Skills Index* was validated in 3 samples of 500 Australian adults, ages 20-44 years. In addition, McKenzie's checklist for evaluating health information [49], which is not an instrument but a decision aid intended for health educators to share with consumers, was used to further support the choice of particular domains of skills to test. The internal consistency of the survey, measured by Cronbach α, was 0.62 considered acceptable [50].

Science Literacy (Knowledge About Science) Survey

This survey, which comprised 12 multiple-choice questions, was developed based on TOSLS, with a number of modifications. TOSLS, which measures the scientific literacy skills of students in undergraduate biology classes, includes questions that assess competency in 9 scientific literacy skills. Of these 9, we adopted 2, *identifying a valid scientific argument* and *understand elements of research design and how they impact scientific findings or conclusions*, as highly relevant to evaluating the content of web-based health claims. We excluded skills that pertained to conducting first-order research inquiry and analyzing and presenting quantitative data (*solve problems using quantitative skills, including probability and statistics*), as we saw these as more relevant to performance in undergraduate biology classes than to consumer health tasks. We also excluded *evaluate the validity of sources*, viewing this as a key information literacy competency, measured by our information literacy test. Finally, we excluded questions that involved judging the appropriateness of the *use of science by government, industry, and media*.

To develop our measure, we adopted the structure and content of some questions and answer options by Gormally et al [34], while simplifying the language (as our primary audience was not a college class) and editing the content to make it applicable to the health domain. We also wrote several additional health-specific questions and answer options that pertained to the 2 scientific literacy skills relevant to our task. The resulting measure comprised 12 questions, 5 assessing the ability to identify a valid scientific argument, and 7 assessing understanding of research design and how it pertains to scientific findings.

TOSLS was developed through an iterative process using built-in validation procedures. According to Gormally et al [34]:

> *...measures of validity included correspondence between items and scientific literacy goals of the National Research Council and Project 2061, findings from a survey of biology faculty, expert biology educator reviews, student interviews, and statistical analyses.*

Although TOSLS validity data do not extend to our measure, TOSLS provides a theoretical foundation for our conceptualization of scientific literacy as it pertains to evaluating web-based health information. The items were developed through several iterative review and discussion rounds by the 4 authors. The internal consistency of our science literacy survey, measured by Cronbach α, was 0.6, which is considered acceptable [50].

Immune System Knowledge Survey

We devised a five-question immunology knowledge assessment test to measure the basic knowledge of the immune system. The questions were developed with the expectation that anyone who had taken a middle or high school biology course could answer them correctly. The issue is whether someone with basic knowledge of the immune system would assess the claims expressed in the video from a more critical vantage point.

Immunology is an immensely complex topic that addresses issues well beyond the scope of our study.

Interpersonal Trust Survey

This survey assessed interpersonal trust and included one question commonly used in interpersonal trust surveys [51]. The question asked, "Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people?" The answers required choosing a number on the scale from 1 to 5, where 1 meant *you can't be too careful* and 5 *most people can be trusted*.

Trust in Health Authority Survey

This survey, developed by the researchers, assessed participants' trust in 5 established authoritative health information sources: the Centers for Disease Control and Prevention, National Institutes of Health, major research universities, national voluntary health associations (the survey included American Diabetes Association as an example), and responders' primary health care providers. The answer options ranged on a 5-point Likert scale, from *don't trust at all* to *trust completely*. The questions were developed based on a literature review of trust in health care and biomedicine [52]. The internal consistency of the health authority trust items, measured by Cronbach α, was 0.64 and considered acceptable [50].

Data Collection and Preparation

The survey data were collected using Qualtrics XM, a widely used survey software. As a quality control measure, we removed responses that were submitted multiple times (those containing near-identical narrative answers) and responses that were completed in less than 20 minutes. The cutoff was determined empirically by the authors by completing the task as rapidly as possible while viewing the entire video and giving meaningful answers. The resulting data set included 150 responses. One of the participants completed the main information evaluation task only; another completed all the tasks except for the scientific literacy survey, which the participant left partially incomplete. The remaining 148 responses were complete. Missing data for the 2 partial responses were imputed from the means of the corresponding variables.

Data Analysis

For quantitative analysis, we used multiple-choice data that did not require coding. Scores for each survey were obtained as a simple count of the correct answers. Open-ended responses were reviewed carefully and used to provide illustrations and outline the narrative context of the quantitative data. SPSS (IBM) was used for quantitative data analysis. All subsections of the Results section, except for the last one, titled "Relationship between willingness to share the video and information competencies," report descriptive statistics illustrated by narrative examples. The last subsection reports the results of the inferential analysis using the statistical methodology described in this subsection.

Results

Willingness to Share the Video and Conceptualization of Evidence

Overall, the participants found the video worth sharing. Of the 150 participants, 70% (105/150) were *very likely* (47/150, 31.3%) or *somewhat likely* (58/150, 38.7%) to recommend the video to their respective cousin, whereas only 11.3% (17/150) were *very unlikely* to do that. Those likely to recommend the video put forward several reasons why they found it worth forwarding. Some had to do with their previous beliefs about the effectiveness of dietary supplements and the particular ingredients in the advertised supplement to enhance the immune system.

For example, one participant stated:

> *I am a firm believer in supplements.*

and another wrote:

> *With ingredients like Zinc and Elderberry, I feel more confident in this pill, as I bombarded myself with elderberry last flu season and it seemed to stop my cold.*

Other reasons had to do with the confidence in the wisdom of the crowd and the quality control afforded by the visibility of public opinion in the internet era, "If the quality is not good, many people [would] have complained." Some expressed trust in the channel or TV show on which the video aired, and with which they had prior familiarity. A number felt that there was no harm in trying something that may potentially be of help (eg, "Since [he] hasn't recovered for a few weeks, try all of them, maybe it works"). However, many others provided statements that did not include justifications (eg, "It's a good choice and I think it will work").

The participants who were unlikely to recommend the video also provided different justifications. Some mentioned the absence of *evidence* of the supplement's effectiveness or the lack of Food and Drug Administration's approval. Many were taken aback by the proposed schedule of taking the supplement every hour. Typically, regardless of their willingness to share the video, participants did not provide multiple reasons and did not weigh *pro* and *con* arguments against one another.

Participants were also presented with a number of hypothetical statements that could be viewed as evidence and asked whether these would affect their opinion about the supplement, either positively or negatively (Table 2 contains statements and the results). The large majority responded that their opinion about the supplement would be positively influenced by factors that do not meet the criteria of valid scientific evidence. These included a friend's account of a positive experience with the supplement (124/150, 82.7% of individuals), high customer ratings on a website of crowdsourced reviews (108/150, 72% of individuals), a survey of supplements consumers done without controls (99/150, 66% of individuals), and a statement on the supplement producer's site that 9 out of 10 people found the product beneficial (98/150, 65.3% of individuals).

Table 2. Likelihood of the following things affecting the opinion about the supplement (N=150).

Statement	Participants, n (%)		
	Positive effect[a]	No effect	Negative effect[a]
You speak to another friend, and she says that she has been taking this product for the past 2 years and has never caught a cold.[b]	124 (82.6)	21 (14)	5 (3.3)
A crowdsourcing review website focused on supplements found that almost every reviewer had positive things to say about the supplement (its average rating being 4.5 out of 5 stars).[b]	108 (72)	29 (19.3)	13 (8.6)
A crowdsourcing review website focused on supplements found that many people were dissatisfied with the supplement (its average rating being 2.3 out of 5 stars).[c]	24 (16)	40 (26.6)	86 (57.3)
A survey of consumers who had been using the product for the last year indicates that 85% of them did not get the flu last year and 15% got the flu.[b]	99 (66)	47 (31.3)	4 (2.6)
On the supplements company's website, they state that a study found that 9 out of 10 people found the product to be beneficial.[b]	98 (65.3)	48 (32)	4 (2.6)
On the supplements company's website, a video explains the benefits of the supplement in terms of the biochemistry of how it boosts the immune system.[b]	92 (61.3)	52 (34.6)	6 (4)
This supplement has never been tested in a controlled clinical trial.[c]	20 (13.3)	31 (20.6)	99 (66)
A year later, a review by scientists from the National Institute of Allergy and Infectious Diseases concludes that scientific evidence about the effectiveness of this supplement is inconclusive.[c]	40 (26.6)	37 (24.6)	73 (48.6)
Another year later, a new review by scientists from the National Institute of Allergy and Infectious Diseases concludes that scientific evidence supports the claim that this supplement reduces frequency (how often) and severity (how bad) of respiratory (breathing-related) infections.[b]	127 (84.6)	18 (12)	5 (3.3)

[a]Combines *very likely* and *somewhat likely*.

[b]May be perceived as *in favor* evidence.

[c]May be perceived as *against* evidence.

In explaining their responsiveness to potential influences, many participants stressed the value of their friends' personal experiences. They described relationships with friends as built on trust, which translated into the information being seen as trustworthy. For example, one participant said, "Personal experience from a trusted person is persuasive." The term *trust*, mentioned frequently, was described as an assurance of accuracy and objectivity:

> *I trust what my friends say. So her experience would make me believe in the product's efficacy more than the information from the morning show guest. I perceive the guest to have a pecuniary interest in sales of the product.*

Many respondents also had positive views of reviews on crowdsourcing sites. Although some mentioned possible reasons for skepticism, such as the number of reviews or reviews by bots, many felt reassured by the absence of negative reviews. For example, one respondent wrote, "So many people use it and the effect is very good, so I believe this product has good quality." However, others preferred to obtain recommendations from their doctors.

In considering a survey that found that 85% of the product's users did not get the flu in a year, most felt reassured by the number, stating that, in their experience, the number was high enough to suggest that the product was effective. There were also some doubts voiced by participants; for example, "I have a feeling that people who seek out and spend money on supplements like these tend to be healthier in many areas of their lives, so are less likely to get the flu in general." Responses to testimonies and claims of effectiveness on the company's site also included a mix of positivity and skepticism. They ranged from "They did research and the results are good, so I think the product quality is good" to "The credibility of research done by his own company is not guaranteed."

Participants also reacted to the hypothetical scientific evidence. A total of 73 (48.6%) participants said their opinion of the supplement would be negatively affected if a study by the NIAID deemed the evidence of the supplement's effectiveness inconclusive. Overall, 99 (66%) participants reacted to "This supplement has never been tested in a controlled clinical trial" by stating that this would affect their opinion *very* or *somewhat* negatively. Had a NIAID study found evidence of the supplement's effectiveness, 127 (84.6%) respondents' opinions would be positively affected. Overall, many participants expressed a view of science as an important foundation of health-related knowledge. They referred to scientists as having "authority" and "credibility." One person stated, "Dr. Fauci and the National Institute of Allergy and Infectious Diseases conduct research in a scientific manner and this is very good." Many expressed trust in NIAID and "national institutions," "public health agencies," and "government agencies." Skeptical voices also existed: "I would still be skeptical, and would want to know about who funded the study, what they found, how much the supplement actually reduced frequency, severity, etc." Yet others explicitly preferred their own conclusions and experience.

Information Competencies

Information Literacy

For the information literacy measure with a possible range of scores from 0 to 6, participants' mean score, or the number of correct responses to the multiple-choice questions, was 3.23 (SD 1.72). Table 3 shows the distribution of information literacy scores.

Responses to individual information literacy questions are summarized in Table 4. In evaluating the reliability of information sources, most participants recognized the value of dot-gov and, to a lesser extent, dot-edu domains. For example, in choosing among 3 possible websites in search of "unbiased information" about food to support a child's immune system, 72.4% (108/149) responders chose a dot-gov site that was "checked by health professionals." In deciding which site was most likely to provide accurate health information, 59.7% (89/149) chose "an institute run by the US government," followed by "a support group for patients living with a particular illness" (46/149, 30.9%). In choosing the most reliable site between Harvard Health Publishing website, Healthline website, Medlinx website, and BBC website, a little over half (77/149, 51.7% of respondents) wrote that the Harvard site was the most reliable. When it came to evaluating author credentials as markers of their authority and qualification to write about the immune system, a little over half (76/149, 51%) participants chose an allergist (a physician specializing in immune deficiency disorders) over a food chemist (28/149, 18.8%), a naturopath (27/149, 18.1%), or a health blogger (18/149, 12.1%).

Table 3. Distribution of participants' information literacy scores (N=149).

Score	Participants, n (%)
0	6 (4)
1	19 (12.8)
2	28 (18.8)
3	39 (26.2)
4	18 (12.1)
5	16 (10.7)
6	23 (15.4)

Table 4. Information literacy, correct responses (N=149).

Question	Answer options[a]	Correct responses, n (%)
Lisa has a toddler and is looking for a website with unbiased information about food to support her daughter's immune system. She finds 3 websites. Which of the sites is the best option for Lisa?	• The 3 options differ in domains (dot-com, *dot-gov*, dot-info, recency, and background of authors)	108 (72.5)
Which of the following sources' websites is most likely to provide accurate health information?	• Four options ranging from a *US federal institution* to a social media company selling a health app	89 (59.7)
You want to find more information about making your immune system stronger. You type *boost immune system* into Google. From the results of that search, which website is likely the most reliable information source?	• *www.health.harvard.edu* • www.healthline.com • www.medlinx.com • www.bbc.com	77 (51.7)
Which of the following authors would be the best qualified to write an article about the immune system?	• a food chemist, PhD • a naturopath, NMD • a food blogger, MA • *an allergist, MD*	76 (51)
You go to the mercola.com website, which features health news and articles. The website states that "The entire contents of this website are based upon the opinions of Dr. Mercola, unless otherwise noted." Does this mean the content has been reviewed by independent medical professionals (eg, qualified doctors, nurses, or other health care providers?)	• Yes, definitely • *No, not necessarily*	72 (48.3)
The developers of a drink called BoostRx claim that their product increases the effectiveness of the immune system. Which of the additional information below would provide the strongest evidence supporting this claim?	• Four options that include *published independent studies*, studies by the developer, advertisements, and purchasers' reviews	60 (40.3)

[a]Correct answers in italics.

Another information literacy question asked participants to determine whether the content of a health information site was verified by independent reviewers. Despite the explicit statement in the question that "the entire contents" were based on the opinions of the site's owner, Dr. Mercola, fewer than half (72/149, 48.3%) of the participants chose *no, not necessarily*, whereas the majority (77/149, 51.7%) believed that the site was *probably* verified by independent medical professionals.

The most difficult information literacy question turned out to be the one asking participants to choose a piece of information that would provide the strongest evidence for the claim that a drink called BoostRx increased the effectiveness of the immune system. Only 40.3% (60/149) of respondents correctly chose links to published studies that did not involve developers. A total of 31.2% (47/149) chose reviews by satisfied purchasers, 24.9% (37/149) preferred articles written by one of the developers, and 5 selected advertisements on the developer's site.

Immune System Knowledge

For the measure of immune system knowledge, with possible scores ranging from 0 to 5, the mean score was 2.38 (SD 1.34). The distribution of scores is presented in Table 5.

The question answered correctly by most of the participants was "What do vaccines do?" The majority, 70.5% (105/149), selected "Stimulate the immune system to produce antibodies." The majority (100/149, 67.1% of participants) also recognized that white blood cells produced antibodies. A little fewer than half (72/149, 48.3%) were able to answer what constituted the first line of defense against microbes (skin). About a third, approximately 36.2% (54/149) correctly selected the answer that described the relationship between antibodies and antigens. Finally, only 16.1% (24/149) recognized that all 3 organs, cells, and chemicals were components of the immune system.

Table 5. Distribution of participants' immune system knowledge scores (N=149).

Score	Participants, n (%)
0	12 (8.1)
1	29 (19.5)
2	38 (25.5)
3	43 (28.9)
4	17 (11.4)
5	11 (7.4)

Science Literacy

For the measure of science literacy, with possible scores ranging from 0 to 12, the mean score was 6.57 (SD 2.40). The distribution of the scores is shown in Table 6.

Table 6. Distribution of participants' science literacy scores (N=148).

Score	Participants, n (%)
0	0 (0)
1	0 (0)
2	0 (0)
3	10 (6.8)
4	23 (15.5)
5	29 (19.6)
6	17 (11.5)
7	22 (14.9)
8	13 (8.8)
9	8 (5.4)
10	14 (9.5)
11	11 (7.4)
12	1 (0.7)

The questionnaire included 2 question formats. The first presented an accurate or faulty reasoning statement and asked, "Is this a good scientific argument?" The other format described the design of a study, either well-designed or problematic, and

a conclusion, and asked whether, based on the presented information only, any other factors could explain the difference (Table 7). In both cases, participants were less likely to recognize faulty reasoning or design than they were to doubt good logic and well-designed experiments. The most difficult questions turned out to be those that required recognizing potential confounding factors in the research design. Considering base rates and sample sizes also proved challenging (see Table 7 for examples).

Table 7. Selected questions requiring recognizing potential confounds in research design (N=148).

Question	Answer[a]	Correct, n (%)
Researchers want to study how noise affects task performance. They randomly put participants into 2 groups. Females make up 35% of the first group and 75% of the second group. Participants in the first group complete a moderately difficult task in a quiet room. Participants in the other group do the same task in a noisy room. Researchers say that any differences in performance between the groups will be because of the noise. Based on this information only, do you see any other factors that may explain the difference?	• *Yes* • No	47 (31.8)
This year, there were 100,000 more cases of adolescent depression diagnosed in the US than last year. Thus, adolescent depression in the US is on the rise. Is this a good scientific argument?	• Yes • *No*	63 (42.6)
Researchers in a cancer clinic test a new drug in 12 patients with a rare cancer. None of the patients experience any dangerous side effects. The researchers concluded that the drug is safe. Based on this information only, do you see any factors in the design that make you less confident about the researchers' interpretation of their findings?	• *Yes* • No	58 (39.2)
Many people who take multi-vitamins do not catch colds frequently. Thus, taking multi-vitamins prevents colds. Is this a good scientific argument?	• Yes • *No*	77 (52)

[a]Correct answers in italics.

Trust

On the measure of interpersonal trust, deciding whether *most people can be trusted* where 1 meant *you can't be too careful* and 5 *most people can be trusted*, the majority of respondents leaned toward trusting people. Although most participants trusted major biomedical research and policy organizations and the health care establishment, a not so small minority was skeptical of them (Table 8). The mean health authority trust score was 20.08 (SD 2.81) out of 25.

Table 8. Trust in health authority (N=149).

	Participants, n (%)				
	1 (do not trust at all)	2	3	4	5 (trust completely)
Majority of people[a]	11 (7.4)	7 (4.7)	47 (31.5)	68 (45.6)	16 (10.7)
National Institutes of Health	1 (0.7)	11 (7.4)	16 (10.7)	55 (36.9)	66 (44.3)
Centers for Disease Control and Prevention	1 (0.7)	9 (6)	21 (14.1)	68 (45.6)	50 (33.6)
Your primary doctor or health care provider	0 (0)	3 (2)	33 (22.1)	71 (47.7)	42 (28.2)
A national health association, such as American Diabetes Association	1 (0.7)	3 (2)	39 (26.2)	61 (40.9)	45 (30.2)
A major university that conducts biomedical research	3 (2)	5 (3.4)	38 (25.5)	65 (43.6)	38 (25.5)

[a]For this question, answer options are 1=cannot be too careful and 5=can be trusted.

Relationship Between Willingness to Share the Video and Information Competencies

For the inferential analysis, we were interested in the impact of information literacy, science literacy, knowledge of the immune system, interpersonal trust, and trust in health authority on the likelihood of recommending the video.

As a first step in the analysis, we looked at pairwise correlations among the independent variables of interest. The variables were highly correlated (Table 9).

Table 9. Correlations among independent variables (with significance level; N=150).

Independent variable	Information literacy	Science literacy	Immune system knowledge	Interpersonal trust	Health authority trust
Information literacy					
r	1.000	0.505[a]	0.286[a]	−0.390[a]	0.143
P value	N/A[b]	<.001	<.001	<.001	.08
Science literacy					
r	0.505[a]	1.000	0.357[a]	−0.228[a]	0.175[c]
P value	<.001	N/A	<.001	.005	.03
Immune system knowledge					
r	0.286[a]	0.357[a]	1.000	−0.157	0.112
P value	<.001	<.001	N/A	.06	.17
Interpersonal trust					
r	−0.390[a]	−0.228[a]	−0.157	1.000	0.235[a]
P value	<.001	.005	.06	N/A	.004
Health authority trust					
r	0.143	0.175[c]	0.112	0.235[a]	1.000
P value	.08	.03	.17	.004	N/A

[a]Significant at the .001 level (2-tailed).

[b]Not applicable.

[c]Significant at the .05 level (2-tailed).

As interventions often target a single variable, we wanted to compare effect sizes of single predictor models. To do this, we conducted 5 single-predictor regression analyses of the independent variables on the 4-level *likelihood of recommending the video*. To correct for multiple hypotheses, the significance of the models was assessed at $P<.01$. The data are summarized in Table 10.

Table 10. Comparing single predictor models.

Independent variable	F test (*df*)	*P* value	Adjusted R^{2a}	Standardized coefficient β	Unstandardized coefficient *B* (CI; SE)	*t* test (*df*)	*P* value
Information literacy	31.24 (1,148)	<.001[b]	0.17	−0.417	−0.236 (−0.320 to −0.153; 0.042)	−5.59 (148)	<.001[b]
Interpersonal trust	29.04 (1,148)	<.001[b]	0.16	0.405	0.393 (0.249 to 0.538; 0.073)	5.38 (148)	<.001[b]
Science literacy	11.65 (1,148)	<.001[b]	0.07	−0.270	−0.110 (−0.174 to −0.046; 0.032)	−3.41 (148)	<.001[b]
Immune system knowledge	6.17 (1,148)	.01[c]	0.03	−0.200	−0.146 (−0.262 to −0.030; 0.059)	−2.48 (148)	.01[c]
Health authority trust	2.40 (1,148)	.12	0.01	0.126	0.044 (−0.012 to −0.101; 0.029)	1.55 (148)	.12

[a]Reflects proportion of variance accounted by the model.

[b]Statistically significant at $P<.01$.

[c]Approaches significance at $P<.01$.

The analysis shows that, as single predictors, information literacy had the largest effect size, predicting the largest amount of variance in the dependent variable (17%). Although only 22% (5/23) of the participants with the highest possible information literacy score of 6 said they would share the video, 95% (18/19) with a score of 1 would share it. The percentage of participants wishing to share the video increased sharply at the information literacy score of 5 (11/16, 69%) and continued to increase (41/57, 72% at the scores of 4 and 3, 25/28, 89% at the score of 2).

The effect size of information literacy was closely followed by that of interpersonal trust (accounting for 16% of the variance). Science literacy as a single predictor accounted for 7% of the variance in the dependent variable and immune system

knowledge for 3%. Trust in health authority had no significant effect as a single predictor.

To compare the strength of associations, we calculated correlations of each predictor with the dependent variable and then used the Fisher r-to-z transformation. The analysis revealed that the top 3 predictors, namely information literacy, interpersonal trust, and science literacy, were not statistically different in the magnitude of their association with the likelihood of sharing the video.

The effects of the variables were in the expected directions: higher information literacy, science literacy, and immune system knowledge scores were associated with a lower likelihood of recommending the video. Higher interpersonal trust was associated with a greater likelihood of recommending the video.

Due to multiple correlations leading to expectation of shared variance, we also performed a linear regression with all 5 independent variables in the single model. Multicollinearity diagnostic tests were performed and did not raise concerns about violations of assumptions for regression (all VIFs≤1.48). The model was statistically significant, $F_{1,144}$=10.60; P<.001, and accounted for 24% of the variance in the likelihood of recommending the video. In the final model with all the predictors entered, information literacy (2-tailed t_{149}=−3.26; P=.001) and interpersonal trust (t_{149}=2.85; P=.005) were statistically significant; trust in health authority approached significance (t_{149}=1.72; P=.09); immune system knowledge and science literacy were not significant with the other factors in the model.

The overall analysis suggests that information literacy, science literacy, immune system knowledge, and interpersonal trust were related to participants' willingness to share the video; the impact of trust in health authority was marginal. Moreover, the correlation among the predictors and different significance of the variables in the single predictor and multiple predictor models suggest the possibility of mediation. For example, it is possible that information literacy mediated the impact of scientific literacy variables, with science literacy variables influencing judgment not directly, but by affecting information literacy.

Discussion

Overview

This study confirms that people are vulnerable to web-based misinformation and are likely to propagate it by sharing it with others [8-19]. In evaluating claims, they are often influenced by the information that science and information professionals within the normative health care paradigm do not consider supporting evidence. Examples include hearsay, majority opinions, or statements by those with conflicts of interest. The study also shows that information literacy, science literacy, and health domain knowledge are challenging competencies. None of this is surprising.

Theoretical and Practical Contributions

The study's major theoretical and practical contribution to the field is in demonstrating that greater information literacy and science literacy are associated with lesser vulnerability to misinformation and lesser propensity to share it. Although the field of eHealth recognizes these literacies, they have received little attention in empirical research [53]. Although the relationship is complex, and vulnerability to misinformation is affected by a host of interacting factors, programs that target information literacy and understanding of science hold promise for helping individuals and communities. Of these 2 factors, information literacy is easier to address in informal educational settings, such as libraries, community organizations, and health clinics. Science literacy, on the other hand, is primarily developed over a period of years in a science classroom. Still, the two correlate and likely influence one another, pointing to the value of a collaborative conversation among school science educators, information professionals, and health professionals.

Another contribution of this study to the field is providing a starting point for building instruments for assessing information literacy and science literacy as they pertain to health information and information behaviors. The instruments developed for this study build on existing tools developed for related purposes. Although developing robust psychometric instruments requires multiple rounds of item testing and adjustment that are beyond the scope of this study, the internal consistencies of our initial instruments pose them as a feasible foundation. Further work is needed to sharpen and validate them for use in consumer health information contexts.

Finally, unlike much research into the public's reaction to web-based health information [8,10,14], this study focuses on a YouTube video, rather than web text. YouTube videos as an information source pose a unique challenge to consumers: their authorship and ownership are often harder to establish, whereas auditory information and the absence of hyperlinks make verifying authority and fact checking more challenging. In addition, existing web-based information evaluation criteria have not been optimized for assessing the quality of videos [54,55]. At the same time, videos are an attractive format to watch and easy to share, and studying consumer behavior in sharing online health videos is important.

The Issue of Trust

This discussion would be incomplete without addressing the issue of trust and the need for further research into the relationship between trust and vulnerability to health misinformation. One important area of trust that merits further investigation is trust in science [52]. In designing our science literacy instrument, we chose to focus on knowledge about science (eg, understanding scientific methods and argumentation), rather than on trust in science as an enterprise. In evaluating potential supporting evidence for the claims made in the video, our participants made many positive claims about science. For example, clinical trials of the effectiveness of the supplement were the type of evidence that raised confidence in the supplement for the greatest number of participants. In future research, we intend to explore trust in science more closely.

The 2 kinds of trust that were investigated in this study, interpersonal trust and trust in health authority, also merit further investigation. It was unexpected that trust in health authority was only marginally associated with vulnerability to misinformation and in only one of the models. This may be due to the specific health information domain addressed by the study, dietary supplements for strengthening the immune system. Scientific data addressing the impact of supplements on health are very complex and do not easily translate into evidence-based public health messaging. In other areas, such as attitudes toward vaccination, trust in health authority and vulnerability to misinformation may have a stronger association.

Interpersonal trust's association with vulnerability to health misinformation posits a philosophical challenge. In a way, it is not surprising that people who are less trusting in general are also more skeptical of overgeneralized, inflated promises of cures, unsupported by research evidence. At the same time, social science research does not equate trust with gullibility [46] and generally views higher interpersonal trust as a characteristic of healthy functional societies. Institutional trust also positively correlates with trust in societal institutions [45] and should be expected to correlate with trust in science and health care establishments. To the best of our knowledge, no studies have conducted an in-depth investigation of different types of trust related to the spread of misinformation on the internet. The topic, however, is important, especially as it relates to misinformation around COVID-19 infodemic, and merits further studies.

Implications for Education and Consumer Health Informatics

Efforts to help the public navigate and share health information in the era of social media can be of 2 kinds. The first kind involves programs, activities, and materials that target individuals' information competencies. Such programs can occur in a variety of settings, focusing on helping individuals analyze the characteristics of information sources, presentation, and content. Our study suggests that such interventions have promise. At the same time, information and science literacy develop over time, growing with many educational and personal experiences, so easy fixes are unlikely. Moreover, these competencies do not tell the full story when it comes to predicting information behavior. As an illustration, in our sample, many participants with high competencies scores expressed an intent to share the video. The science literacy background section of this paper discusses how people often make choices that appear inconsistent with their scientific knowledge, driven by considerations of group identity [38] or fast reactive responses [40].

The second kind of approach to supporting the public involves leveraging technologies. Examples of technologies can vary, from fact-checking sites to machine learning misinformation detection tools [56]. The development of such tools is important because it alleviates the burden of being vigilant for consumers and also reduces the cognitive load while processing complex information. This entails less time commitment than educational efforts.

However, technology alone will not solve the problem of misinformation. The challenge of evaluating information authority applies to a fact-checking tool as much as it applies to a YouTube video selling dietary supplements. Moreover, trust in information and information providers is critical for the acceptance of any information technology. Helping individuals recognize web-based health misinformation is a complex, multifaceted enterprise that should involve the collaborative efforts of researchers, technology developers, librarians, educators, and community outreach specialists.

Limitations and Directions for Future Research

As this study was conducted with a small nonrepresentative sample, it does not provide information on the general picture of the propensity to share health misinformation on the internet, as well as related literacies, in the population. Moreover, the study used a hypothetical scenario and could not draw on the complex social factors that affect decisions to share information in real life (eg, approval and status seeking). Supplementing this study with research into real-world health information sharing decision making will provide more nuanced data for future understanding of the issue.

Furthermore, no validated psychometric instruments exist currently for assessing information literacy and science literacy in the context of web-based health information behavior. Although our research instruments were designed with care, the formal process of designing, testing, adjusting, and validated psychometric tools was outside the scope of our study. Although the internal consistency data of our instruments were acceptable, a higher reliability is expected of formal instruments. We hope that as attention to information literacy, science literacy, and trust grows, such robust instruments will emerge. At present, our results should be interpreted with the understanding that they are affected by our specific selection of questions and answer options in our instruments. As the lower internal consistency of instruments makes the effects more difficult to detect, the sizes of the effects found here may be greater in future studies.

Finally, our study involved a survey methodology with a primarily quantitative focus, an approach that works better for establishing relationships than for describing them in depth. An interview study probing participants' reasoning would provide a nuanced richness of information on this important topic.

Acknowledgments

This work was supported in part by the intramural research program at the National Library of Medicine, National Institutes of Health. The authors thank Yulia Chentsova Dutton for sharing her invaluable statistical expertise. The authors are also thankful to the Network of the National Library of Medicine for helping with inviting participants to the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
All instruments.
[DOCX File , 26 KB - jmir_v23i12e30323_app1.docx]

References

1. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
2. Cuan-Baltazar JY, 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]
3. Diviani N, van den Putte B, Meppelink CS, van Weert JC. 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]
4. Kim HK, Ahn J, Atkinson L, Kahlor LA. Effects of COVID-19 misinformation on information seeking, avoidance, and processing: a multicountry comparative study. Sc Commun 2020 Sep 13;42(5):586-615. [doi: 10.1177/1075547020959670]
5. Molina MD, Sundar SS, Le T, Lee D. "Fake News" is not simply false information: a concept explication and taxonomy of online content. Am Behav Sci 2019 Oct 14;65(2):180-212. [doi: 10.1177/0002764219878224]
6. Wardle CH, Derakhshan H. Information Disorder: Toward An Interdisciplinary Framework for Research and Policy Making. France: Council of Europe; 2017.
7. Keselman A, Arnott Smith C, Murcko AC, Kaufman DR. Evaluating the quality of health information in a changing digital ecosystem. J Med Internet Res 2019 Feb 08;21(2):e11129 [FREE Full text] [doi: 10.2196/11129] [Medline: 30735144]
8. Sommariva S, Vamos C, Mantzarlis A, Đào LU, Martinez Tyson D. Spreading the (Fake) news: exploring health messages on social media and the implications for health professionals using a case study. Am J Health Educ 2018 Jun 07;49(4):246-255. [doi: 10.1080/19325037.2018.1473178]
9. Wells G. Twitter to Add Labels to Disputed Coronavirus Posts, as Misinformation Proliferates. Wall Street Journal. 2020. URL: https://www.wsj.com/articles/twitter-to-add-labels-to-disputed-coronavirus-posts-as-misinformation-proliferates-11589226911 [accessed 2021-09-21]
10. Nsoesie EO, Cesare N, Müller M, Ozonoff A. COVID-19 misinformation spread in eight countries: exponential growth modeling study. J Med Internet Res 2020 Dec 15;22(12):e24425 [FREE Full text] [doi: 10.2196/24425] [Medline: 33264102]
11. 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]
12. Shin J, Valente T. Algorithms and health misinformation: a case study of vaccine books on Amazon. J Health Commun 2020 May 03;25(5):394-401. [doi: 10.1080/10810730.2020.1776423] [Medline: 32536257]
13. Ferrand J, Hockensmith R, Houghton RF, Walsh-Buhi ER. Evaluating smart assistant responses for accuracy and misinformation regarding human papillomavirus vaccination: content analysis study. J Med Internet Res 2020 Aug 03;22(8):e19018 [FREE Full text] [doi: 10.2196/19018] [Medline: 32744508]
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. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? Child Care Health Dev 2017 Jul;43(4):499-503. [doi: 10.1111/cch.12442] [Medline: 28105642]
16. Leong AY, Sanghera R, Jhajj J, Desai N, Jammu BS, Makowsky MJ. Is YouTube useful as a source of health information for adults with type 2 diabetes? A South Asian perspective. Can J Diabetes 2018 Aug;42(4):395-403.e4. [doi: 10.1016/j.jcjd.2017.10.056] [Medline: 29282200]
17. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: a quantitative analysis of YouTube videos. Hum Vaccin Immunother 2018 Jul 03;14(7):1654-1659. [doi: 10.1080/21645515.2018.1454572] [Medline: 29553872]
18. Krishna A. Motivation with misinformation: conceptualizing lacuna individuals and publics as knowledge-deficient, issue-negative activists. J Public Relat Res 2017 Sep 11;29(4):176-193. [doi: 10.1080/1062726x.2017.1363047]
19. Chen L, Wang X, Peng T. Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]
20. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JY, 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]

21. Institute of Medicine Committee on Health Literacy. Health Literacy: A Prescription to End Confusion. Washington (DC): National Academies Press (US); 2004.
22. Logan R. Health literacy research's growth, challenges, and frontiers. In: Arnott Smith C, Keselman A, editors. Meeting Health Information Needs Outside of Healthcare: Opportunities and Challenges. Oxford, UK: Chandos; 2015:19-38.
23. Keselman A. Health Literacy and Other Competencies: The Skills Consumers Need in Order to Be Effective in the Digital Health Information Environment. In: Arnott Smith C, Keselman A, editors. Consumer Health Informatics: Enabling Digital Health for Everyone. London, UK: Chapman and Hall/CRC; 2020:33-56.
24. Pinto M, Cordón JA, Gómez Díaz R. Thirty years of information literacy (1977—2007): a terminological, conceptual and statistical analysis. J Libr Inf Sci 2010 Jan 06;42(1):3-19. [doi: 10.1177/0961000609345091]
25. Association of American Colleges and Universities. The LEAP vision for learning: Outcomes, practices, impact, and employers' views. Washington, DC: Association of American Colleges and Universities; 2011.
26. Presidential Committee on Information Literacy: Final Report. Association of American College and Research. 1989. URL: https://www.ala.org/acrl/publications/whitepapers/presidential [accessed 2021-09-21]
27. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010 Aug 31;15 Suppl 2(sup2):9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
28. Birdsong L, Freitas J. Helping the non-scholar scholar: information literacy for lifelong learners. Library Trends 2012;60(3):588-610. [doi: 10.1353/lib.2012.0003]
29. Bybee R, McCrae B, Laurie R. PISA 2006: an assessment of scientific literacy. J Res Sci Teach 2009 Oct;46(8):865-883. [doi: 10.1002/tea.20333]
30. Layton D, Jenkins E, Macgill S, Davey A. Inarticulate Science? Perspectives on the Public Understanding of Science and Some Implication for Science Education. East Yorkshire, UK: Studies in Education Ltd of Driffield; 1993.
31. Keselman A, Kaufman DR, Patel VL. "You can exercise your way out of HIV?" and other stories: the role of biological knowledge in adolescents' evaluation of myths. Sci Ed 2004 Jul;88(4):548-573. [doi: 10.1002/sce.10135]
32. Feinstein N. Salvaging science literacy. Sci Ed 2010 Sep 20;95(1):168-185 [FREE Full text] [doi: 10.1002/sce.20414]
33. Zeyer A. A win-win situation for healthscience education: seeing through the lens of a new framework model of health literacy. In: Zeyer A, Kyburz-Graber R, editors. Environment | Health: Towards a renewed pedagogy for science education. Netherlands: Springer; 2012:147-173.
34. Gormally C, Brickman P, Lutz M. Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. CBE Life Sci Educ 2012;11(4):364-377 [FREE Full text] [doi: 10.1187/cbe.12-03-0026] [Medline: 23222832]
35. Agley J, Xiao Y. Misinformation about COVID-19: evidence for differential latent profiles and a strong association with trust in science. BMC Public Health 2021 Jan 07;21(1):89 [FREE Full text] [doi: 10.1186/s12889-020-10103-x] [Medline: 33413219]
36. Jennings W, Stoker G, Willis H, Valgardsson V, Gaskell J, Devine D, et al. Lack of trust and social media echo chambers predict COVID-19 vaccine hesitancy. medRxiv. Preprint posted online January 27, 2021 [FREE Full text] [doi: 10.1101/2021.01.26.21250246]
37. Kahan DM. Misconceptions, misinformation, and the logic of identity-protective cognition. SSRN. 2017. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2973067 [accessed 2021-09-21]
38. Kahan DM, Peters E, Wittlin M, Slovic P, Ouellette LL, Braman D, et al. The polarizing impact of science literacy and numeracy on perceived climate change risks. Nature Clim Change 2012 May 27;2(10):732-735. [doi: 10.1038/nclimate1547]
39. Pennycook G, Cannon TD, Rand DG. Prior exposure increases perceived accuracy of fake news. J Exp Psychol Gen 2018 Dec;147(12):1865-1880 [FREE Full text] [doi: 10.1037/xge0000465] [Medline: 30247057]
40. Pennycook G, Rand DG. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition 2019 Jul;188:39-50. [doi: 10.1016/j.cognition.2018.06.011] [Medline: 29935897]
41. Kahneman D. Thinking, Fast and Slow. New York, NY: Macmillan; 2011.
42. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul 30;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]
43. Hammond K, Hamm R, Grassia J, Pearson T. Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. IEEE Trans Syst Man Cybern 1987 Sep;17(5):753-770. [doi: 10.1109/tsmc.1987.6499282]
44. Carl N, Billari F. Generalized trust and intelligence in the United States. PLoS One 2014 Mar 11;9(3):e91786 [FREE Full text] [doi: 10.1371/journal.pone.0091786] [Medline: 24619035]
45. Giordano GN, Björk J, Lindström M. Social capital and self-rated health--a study of temporal (causal) relationships. Soc Sci Med 2012 Jul;75(2):340-348. [doi: 10.1016/j.socscimed.2012.03.011] [Medline: 22537483]
46. Rotter JB. Interpersonal trust, trustworthiness, and gullibility. Am Psychol 1980;35(1):1-7. [doi: 10.1037/0003-066x.35.1.1]
47. TMJ4 News. Supplement to boost your immune system. 2019. URL: https://www.youtube.com/watch?v=BAnkPoV4ioU [accessed 2021-09-15]

48. Ayre J, Costa DS, McCaffery KJ, Nutbeam D, Muscat DM. Validation of an Australian parenting health literacy skills instrument: the parenting plus skills index. Patient Educ Couns 2020 Jun;103(6):1245-1251. [doi: 10.1016/j.pec.2020.01.012] [Medline: 31982204]
49. McKenzie JF. A checklist for evaluating health information. J Sch Health 1987 Jan;57(1):31-32. [doi: 10.1111/j.1746-1561.1987.tb00520.x] [Medline: 3643355]
50. Taber K. The use of Cronbach's Alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
51. OECD Guidelines on Measuring Trust. Paris: OECD Publishing; Nov 23, 2017.
52. Key findings about Americans' confidence in science and their views on scientists' role in society. Pew Research Center. 2020. URL: https://www.pewresearch.org/fact-tank/2020/02/12/key-findings-about-americans-confidence-in-science-and-their-views-on-scientists-role-in-society/ [accessed 2021-09-15]
53. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27]
54. The DISCERN Project 1996-7. DISCERN Online. 2019. URL: http://www.discern.org.uk/background_to_discern.php [accessed 2021-09-30]
55. Boyer C, Baujard V, Nater T, Scherrer J, Appel R. QOI6/416: the health on the net code of conduct for medical and health-related web sites: three years on. J Med Internet Res 1999 Sep 19;1:e99. [doi: 10.2196/jmir.1]
56. Tools that fight disinformation online. Rand Corporation. 2020. URL: https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html [accessed 2021-09-30]

Abbreviations

NIAID: National Institute of Allergy and Infectious Diseases
NNLM: National Network of the Libraries of Medicine
RML: Regional Medical Library
TOSLS: Test of Scientific Literacy Skills

Edited by R Kukafka; submitted 10.05.21; peer-reviewed by Y Zhang, C Urquhart, MA Bahrami; comments to author 28.06.21; revised version received 03.08.21; accepted 12.08.21; published 09.12.21.

Please cite as:
Keselman A, Arnott Smith C, Leroy G, Kaufman DR
Factors Influencing Willingness to Share Health Misinformation Videos on the Internet: Web-Based Survey
J Med Internet Res 2021;23(12):e30323
URL: https://www.jmir.org/2021/12/e30323
doi:10.2196/30323
PMID:34889750

Original Paper

Infodemiological Examination of Personal and Commercial Tweets About Cannabidiol: Term and Sentiment Analysis

Jason Turner[1], MSc; Mehmed Kantardzic[1*], PhD; Rachel Vickers-Smith[2*], PhD

[1]Data Mining Lab, Department of Computer Science and Engineering, University of Louisville, Louisville, KY, United States

[2]Department of Epidemiology, College of Public Health, University of Kentucky, Lexington, KY, United States

[*]these authors contributed equally

Corresponding Author:
Jason Turner, MSc
Data Mining Lab
Department of Computer Science and Engineering
University of Louisville
222 Eastern Pkwy
Louisville, KY, 40292
United States
Phone: 1 502 852 6304
Email: jason.turner@louisville.edu

Abstract

Background: In the absence of official clinical trial information, data from social networks can be used by public health and medical researchers to assess public claims about loosely regulated substances such as cannabidiol (CBD). For example, this can be achieved by comparing the medical conditions targeted by those selling CBD against the medical conditions patients commonly treat with CBD.

Objective: The objective of this study was to provide a framework for public health and medical researchers to use for identifying and analyzing the consumption and marketing of unregulated substances. Specifically, we examined CBD, which is a substance that is often presented to the public as medication despite complete evidence of efficacy and safety.

Methods: We collected 567,850 tweets by searching Twitter with the Tweepy Python package using the terms “CBD” and “cannabidiol.” We trained two binary text classifiers to create two corpora of 167,755 personal use and 143,322 commercial/sales tweets. Using medical, standard, and slang dictionaries, we identified and compared the most frequently occurring medical conditions, symptoms, side effects, body parts, and other substances referenced in both corpora. In addition, to assess popular claims about the efficacy of CBD as a medical treatment circulating on Twitter, we performed sentiment analysis via the VADER (Valence Aware Dictionary for Sentiment Reasoning) model on the personal CBD tweets.

Results: We found references to medically relevant terms that were unique to either personal or commercial CBD tweet classes, as well as medically relevant terms that were common to both classes. When we calculated the average sentiment scores for both personal and commercial CBD tweets referencing at least one of 17 medical conditions/symptoms terms, an overall positive sentiment was observed in both personal and commercial CBD tweets. We observed instances of negative sentiment conveyed in personal CBD tweets referencing autism, whereas CBD was also marketed multiple times as a treatment for autism within commercial tweets.

Conclusions: Our proposed framework provides a tool for public health and medical researchers to analyze the consumption and marketing of unregulated substances on social networks. Our analysis showed that most users of CBD are satisfied with it in regard to the condition that it is being advertised for, with the exception of autism.

(J Med Internet Res 2021;23(12):e27307) doi:10.2196/27307

KEYWORDS

social media; social networks; text mining; CBD; cannabidiol; cannabis; public health; drug regulation; Twitter; sentiment analysis; unregulated substances

Introduction

Although the cannabis plant has been used as a medication for centuries, the use of cannabis was criminalized in the United States in 1937. However, beginning in the 1990s, a few states began allowing the medical use of cannabis even as the plant remained illegal at the federal level [1]. As more states introduced laxer cannabis policies, public interest in the medicinal properties of cannabis evolved to embrace cannabidiol (CBD). CBD is an active chemical found in variants of the cannabis plant that does not have the psychoactive side effects of the tetrahydrocannabinol (THC) components of the cannabis plant [2].

In recent years, the utility of medical cannabis has grown as a major point of discussion in public policy, and especially in online social media discourse [3,4]. In particular, consumers have reported using CBD to treat various conditions, including epilepsy and other neurological disorders, insomnia, and some mental illnesses. CBD remains unregulated by the Food and Drug Administration (FDA) and has not been subjected to the same conclusive trials as most medications pertaining to its many specific uses. In fact, the FDA has only approved one cannabis-derived medication and three cannabis-related drugs to date—all of which require a prescription—and it has not approved marketing cannabis as a safe and effective drug for treating any disease [5]. Palmieri et al [6] reported promising results of CBD as a treatment for inflamed skin conditions and scars. There have also been many studies performed on CBD as a treatment for anxiety and sleep disorders [7-10], as a pain reliever [11-13], and as a treatment for cancer and cancer side effects [13-15].

Although most CBD-based medications and nutritional supplements have not demonstrated safety and efficacy for the numerous indications for which they are used, there are nevertheless many claims in public circulation regarding the effectiveness of CBD for a wide spectrum of ailments.

Social media serves as a primary location for viral CBD marketing and individuals sharing their experiences of personal use. Twitter, in particular, is a useful platform for understanding how cannabis, including CBD, is marketed to consumers and how individuals are using cannabis, as it provides a large corpus of both personal and commercial tweets [3]. Additionally, sentiment analysis can be used on personal and commercial CBD tweets to gauge user satisfaction for CBD treatment of particular medical conditions.

Thus, we propose a framework for the use of text mining in social networks that can help public health experts understand the landscape of personal and commercial claims and sentiments about unregulated substances such as CBD. There are two practical advantages to this framework. First, the data are readily available, easy to access, and inexpensive to use compared to administering surveys or utilizing data from governments and health providers. Second, public health researchers have already shown that sentiment analysis is an effective tool for understanding the public perception of drugs, diseases, and medical services as well as for detecting certain forms of depression [16-19].

To demonstrate the usefulness of this framework, we distinguished between tweets reflecting personal CBD use and tweets reflecting the sales, promotion, and commercialization of CBD. The two resulting corpora of tweets were analyzed for medical-related terms such as conditions, side effects, anatomical terminology, and use of other substances mentioned. This approach allowed us to identify terms that are being used in online CBD marketing in relation to the terms referenced by individuals taking CBD. In addition, we performed sentiment analysis on personal CBD tweets to gauge public opinion on the effectiveness of CBD in treating certain conditions. We were also able to use the Valence Aware Dictionary for Sentiment Reasoning (VADER) model to compute the sentiment of the personal and commercial CBD tweets making reference to specific medical conditions. The VADER sentiment model was used as it was specifically developed for analyzing sentiment on posts made on social network sites such as Twitter [20]. Since its development, VADER has been used to gauge sentiment on digital assistants among Twitter users, political tweets, as well as course evaluations [21-24]

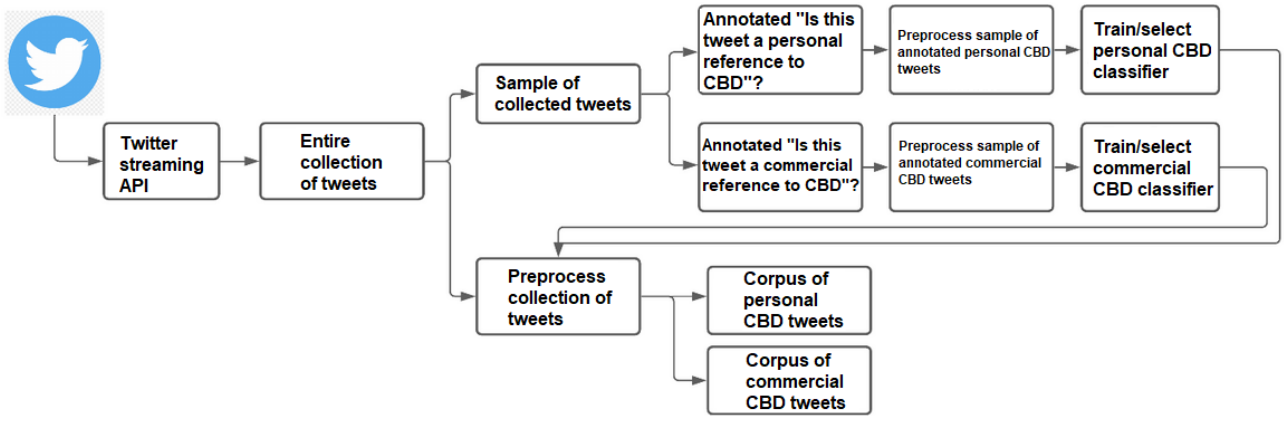
Findings from this study will provide important information on public perceptions of CBD and how CBD is marketed through a social media platform. Misinformation has increased with the rise of social media. Recently, researchers have used social networks to analyze these concepts with regard to medical misinformation [25-28], which can contribute to negative health outcomes. For this reason, unverified claims overstating, exaggerating, and lying about CBD's medical benefits have circulated freely on the internet, leading many to wonder how to assess the actual benefits (if any) of using CBD in medical contexts. Some of the conditions referenced in these claims include ear pain in infants, autism, attention deficit and hyperactivity disorder, Parkinson disease, and Alzheimer disease [29]. By analyzing personal and commercial tweets about CBD using the adaptable and generalizable framework developed for this study, public health and medical professionals can more readily identify viral misinformation pertaining to CBD claims.

Methods

Framework Development

To build this framework, we collected tweets that make reference to "CBD" or "cannabidiol." We then took a random sample of these tweets and labeled them as a personal CBD reference (true/false) or commercial CBD reference (true/false). Using these annotated tweets, we trained two binary text classifiers to separate the personal and commercial CBD tweets from the larger collection of tweets. Using these two corpora of tweets, we identified the most frequently occurring medically relevant terms (terms related to diseases, conditions, symptoms, body parts, other substances, cannabis, etc) and compared these term frequencies between the two corpora. Figure 1 provides a visual description of our framework from data collection through classification. We then computed the sentiment of the personal CBD tweets containing specific terms using the VADER model to assess the satisfaction of CBD for treating those conditions.

Figure 1. Workflow of Tweet collection and classification. API: application programming interface; CBD: cannabidiol.



Data Collection

We collected tweets from the Twitter public stream using the Tweepy Python package as an interface to the live Twitter stream, which provides access to approximately 1% of the public tweets as they are created. Our data collection ran from October 7, 2019, to January 26, 2020, and used the search terms "CBD" and "cannabidiol." We selected this period as it represents the time when CBD had become popular. We restricted our collection to an approximate 3.5-month period so that we could collect a sufficient amount of tweets within a window of time to avoid potential concept drift within the data. We also set filters to collect tweets that were written in English and were original (ie, no retweets). We did not want to include retweets as the actual contents of these tweets are linked to another author and are nearly identical in text to an existing tweet. For each tweet collected, we kept the full-length tweet text, the ID of the tweet, the time the tweet was created, and the Twitter user that authored the tweet. The resulting collection consisted of a data set of 567,850 tweets.

Tweet Annotation

To identify the personal and commercial-related CBD tweets from our collection of 567,850 tweets, we built two binary classifiers trained on a sample of 5496 tweets. This sample of tweets was obtained by taking a 6000-tweet sample from our collection and removing entries with verbatim duplicate tweets. The process of annotating the personal CBD tweets consisted of evaluating each tweet in the sample as to whether or not it was posted from an individual (ie, not a "bot") discussing the past, current, and/or future use of CBD. The process of annotating the commercial CBD tweets consisted of evaluating each tweet in the sample as to whether or not it was posted from an actual (ie, not a "bot") nonnews entity selling, advertising, or promoting CBD. These classifiers were used to distinguish between personal and commercial CBD-related tweets. To train these classifiers, all tweets in the sample were manually labeled as either a personal CBD-related tweet or a nonpersonal CBD-related tweet, and either a commercial CBD-related tweet or a noncommercial CBD-related tweet, according to the content in their full text, which consisted of a maximum of 280 characters. Textbox 1 provides some examples of the personal and commercial CBD-related tweets, and Textbox 2 provides some examples of the types of tweets referencing CBD that we encountered that were considered erroneous for both the personal and commercial CBD classes.

Textbox 1. Examples of personal and commercial cannabidiol (CBD)-related tweets (paraphrased slightly for anonymity). CBC: cannabichromene.

Personal CBD-related tweets.

CBD products are so good for anxiety, and they don't make you high

I've used CBD for anxiety. It is WAY healthier than taking benzodiazepines…I also use CBC for pain. You know what else is bad for your liver? Tylenol and Ibuprofen

Take some painkillers with sleeping aid like Tylenol or Advil PM or something…any CBD or weed maybe try that too

CBD gummies will not give you the high, but for me personally CBD oil edibles helped with anxiety and menstrual cramps

Commercial CBD-related tweets

Go Away!!!!Pain! We have a variety of CBD products for your needs.... Make sure to ask about our selection your next visit. URL

Over time, poor sleep can leave you feeling wrecked...Could CBD help? URL #cbd #cbdoil #hemp #cannabis #sleep #insomnia

Chronic Fatigue...Cannabis CBD THC oil – URL

Our CBD cream combines the relief potential of arnica and natural menthol oil with cocoa butter and the scents of eucalyptus & lavender

Textbox 2. Examples of erroneous tweets (paraphrased slightly for anonymity). CBC: cannabichromene; CBD: cannabidiol; FDA: Food and Drug Administration; THC: tetrahydrocannabinol.

- If you live where medical marijuana is legal, get paid $3k a month to critique weed, CBD, edibles and more URL
- The FDA is worried about CBD. Should you be concerned? URL
- This room is half the size of my cbd apartment...
- Flinders Street in Melbourne's CBD has been re-opened following an earlier protest....Thanks for your patience during this disruption. #victraffic

Analysis of the manually annotated tweets indicated that the classes of personal and commercial CBD-related data sets were imbalanced; the nonpersonal CBD-related tweets occurred 7.7 times more than the personal CBD-related tweets, and the noncommercial CBD-related tweets occurred 10.2 times more than the commercial CBD-related tweets. To achieve a balance of the classes in the training set, we downsampled both of the positive classes in the training set by taking a random sample equivalent in size to the negative class. Table 1 and Table 2 show the class frequencies for the personal and commercial CBD-related tweet classes, respectively, prior to and following downsampling.

Table 1. Training set for personal cannabidiol (CBD) class counts.

Class	Predownsampling, n	Postdownsampling, n
Personal CBD	631	631
Nonpersonal CBD	4865	631
Total	5496	1262

Table 2. Training set for commercial cannabidiol (CBD) class counts.

Class	Predownsampling, n	Postdownsampling, n
Commercial CBD	489	480
Noncommercial CBD	45,007	489
Total	5496	978

Classification Training

Before training the binary classifiers to sort the full data set into personal and commercial CBD-related tweets, we preprocessed the text of the tweets by normalizing all URLs to one consistent string, removing special characters and English part of speech, converting all of the text to lowercase, and lemmatization. This preprocessing was performed to reduce the noise in the data, which would potentially impact the performance of our tweet classifiers. The binary classifiers were then trained on 80% (4396 of 5496) of the annotated sample and validated on the remaining 20% (1099 of 5496) of the annotated sample. We then created a matrix of the term frequency-inverse document frequency (TF-IDF) features based on the words within tweets using a range of n-grams from 1 to 3, as well as a matrix of the TF-IDF features based on the characters within tweets using a range of n-grams from 3 to 6. The resulting matrices were stacked horizontally, which served as the input to our model for training the classifiers.

To train the two binary classifiers, we performed a 5-fold cross-validation grid search using a logistic regression model to find the optimal combination of parameters. The range of parameters is shown in Table 3. After training the binary classifiers, we applied each model to the larger CBD corpora of tweets. To compensate for the small validation set due to the balancing, we performed an additional postclassification test on a random set of 500 unbalanced tweets from our collection to confirm that our models would perform well on real-world unbalanced data. This sample was annotated with the same approach as used for the training set, with the predicted result hidden. We will discuss classification in further detail within the Results section.

Table 3. Parameters used in text classification tuning with the logistic regression model.

Parameter	Range
Penalty	{none,.l1.l2}
Regularization parameter	$x_k=10^{a+(b-a)(k-1)/(n-1)}$, k=1,…n; a=0; b=5; n=20
Solver	{newton-cg, lbfgs, liblinear, sag, saga}

Term Analysis

To track the medical terminology referenced in the sorted commercial and personal CBD tweets, we computed the term frequencies of the top 1000 words in both corpora of tweets. We then confirmed whether these terms were related to relevant medical conditions, medical symptoms, body parts, and/or other

medications/substances by referencing standard English, medical (Systemized Nomenclature of Medicine-Clinical Terms [SNOMED CT]), and slang dictionaries. We categorized the terms into three groups: health/medical, cannabis-related terms, and other substances. Within the health/medical group, we included terms related to diseases, aliments, symptoms, and body parts. We applied the same logic to the terms that appeared to be hashtags by examining the individual words that formed the hashtag for relevancy. We grouped cannabis-related terms together and separated them from the other substances group since there seemed to be an overlap of CBD- and THC-related tweets, both of which referenced the broader cannabis plant; we included cannabis-related slang terms as well as foods that are commonly associated with cannabis infusion (eg, gummies, honey). The other substances group included terms that refer to any other drug or medication. There were a few instances of words being included in multiple groups. For instance, "high" is a side effect of cannabis but is a term commonly used in both cannabis and CBD tweets. Additionally, we considered terms that may represent side effects caused by taking a substance, especially terms commonly associated with cannabis. Finally, we compared the overall frequency of the top occurring terms relative to their frequency in either the personal or commercial class of tweets, and produced a visualization of relevant term frequencies. We used the Scattertext Python package to generate a graphical representation of the frequencies within the personal and commercial CBD classes for each of the three term groups [30].

Sentiment Analysis

We used the VADER model to compute the sentiment of the personal and commercial CBD tweets that reference specific medical conditions. Since this sentiment model incorporates punctuation and text capitalization into computing sentiment, we used the raw tweet text as an input for the model. The VADER model produced a normalized score between –1 and +1 for each tweet based on the summation of the valence scores of each word in the tweet. We converted each tweet's score into a 3-level categorical variable based on the threshold recommended by Hutto and Gilbert [20]: (1) positive sentiment, compound score≥0.05; (2) neutral sentiment, compound score>–0.05 and compound score<0.05; and (3) negative sentiment, compound score ≤–0.05.

We then analyzed the distribution of the compound scores and sentiment class (positive, negative, neutral) in tweets containing terms related to specifically defined conditions and symptoms. Since the VADER model is partially based on a dictionary score and since many terms related to illness may influence the overall sentiment of a tweet (eg, pain, stress, cancer), as part of our analysis, we computed the VADER sentiment scores both with and without medical terms of interest, and compared the mean VADER scores using a *t* test to determine whether any individual term of interest biased the overall sentiment assigned to a tweet. The purpose of the *t* tests was to determine if there was a statistically significant difference in the sentiment conveyed in commercial CBD tweets containing certain terms versus the sentiment conveyed in personal CBD tweets referencing the same terms. For instance, the VADER sentiment score of "CBD really helps my pain" was –0.171 versus a VADER sentiment score of 0.4391 for "CBD really helps my," where the word "pain" holds such a negative VADER sentiment score on its own that it is influencing the overall sentiment score of the tweet.

Ethics

This study leverages publicly available data and is registered as approved by the University of Louisville Institutional Review Board (approved protocol 20.1122).

Results

Classification

We trained the binary classification algorithms for the personal and commercial CBD tweets independently. In both the optimal personal CBD classifier (logistic regression: C=3.36, penalty=none, solver="newton-cg") and commercial CBD classifier (logistic regression=428.13, penalty="l1", solver="saga"), we observed a decrease in classification performance between the smaller validation set derived from balanced data from the unbalanced sample. Despite this decrease in classification performance on the unbalanced data, both the personal and commercial CBD classification models were able to achieve area under receiver operating characteristic curve scores above 0.80. Table 4 and Table 5 show the performance of the personal and commercial CBD binary classifiers, respectively. When the personal CBD binary classifier was applied to the collection of tweets, it classified 167,755 tweets as personal CBD-related tweets. When the commercial CBD binary classifier was applied to the collection of tweets, it classified 143,322 tweets as commercial CBD-related tweets.

Table 4. Personal cannabidiol (CBD) logistic regression classifier performance metrics.

Classification	Precision	Recall	F1	Support	Accuracy	AUC[a]
Balanced sample					0.85	0.86
Nonpersonal CBD	0.93	0.79	0.85	138		
Personal CBD	0.79	0.93	0.85	115		
Unbalanced sample					0.89	0.87
Nonpersonal CBD	0.94	0.91	0.93	367		
Personal CBD	0.78	0.83	0.81	133		

[a]AUC: area under the receiver operating characteristic curve.

Table 5. Commercial cannabidiol (CBD) logistic regression classifier performance metrics.

Classifier	Precision	Recall	F1	Support	Accuracy	AUC[a]
Balanced sample					0.89	0.89
Noncommercial CBD	0.92	0.85	0.89	95		
Commercial CBD	0.87	0.93	0.90	101		
Unbalanced sample					0.87	0.82
Noncommercial CBD	0.90	0.93	0.91	367		
Commercial CBD	0.79	0.70	0.74	133		

[a]AUC: area under the receiver operating characteristic curve.

Term Analysis

We generated unigram frequencies for both the personal and commercial corpora of tweets. We looked at the top 1000 occurring terms (excluding common English stop words) and manually checked if the terms were relevant to health; wellness; diseases; side effects; conditions; body parts; and/or references to other substances against standard English, medical, and slang dictionaries.

In other tweets making cannabis references (Figure 2), it seemed that THC-related terms were mentioned in both the personal and commercial corpora of tweets, with hashtags containing these references more frequently found in the commercial CBD tweets. The terms drink, melatonin, and pills were mentioned in the other substances group (Figure 3) in both the personal and commercial CBD tweets. Kratom and medium-chain triglyceride (MCT) were mentioned more frequently within the commercial CBD tweets and less frequently in the personal CBD tweets. References to alcohol occurred slightly more than average within the personal CBD tweets and below average in the commercial CBD tweets. Opioids were mentioned but only infrequently in both the personal and commercial CBD tweet classes. In the health and wellness group (Figure 4), pain, sleep, and anxiety occurred frequently in both the personal and commercial CBD classes. Terms related to fitness and nutrition were found more frequently in the commercial CBD tweets. Tweets referencing posttraumatic stress disorder (PTSD) occurred at the same average within both classes. Finally, CBD tweets referencing autism occurred more frequently than average in personal tweets, but infrequently in commercial tweets. This is despite the US FDA sending out warning letters to CBD sellers for disseminating misinformation by promoting CBD as a treatment for a variety of medical conditions, including autism [29].

Figure 2. Cannabis-related term frequency per class.

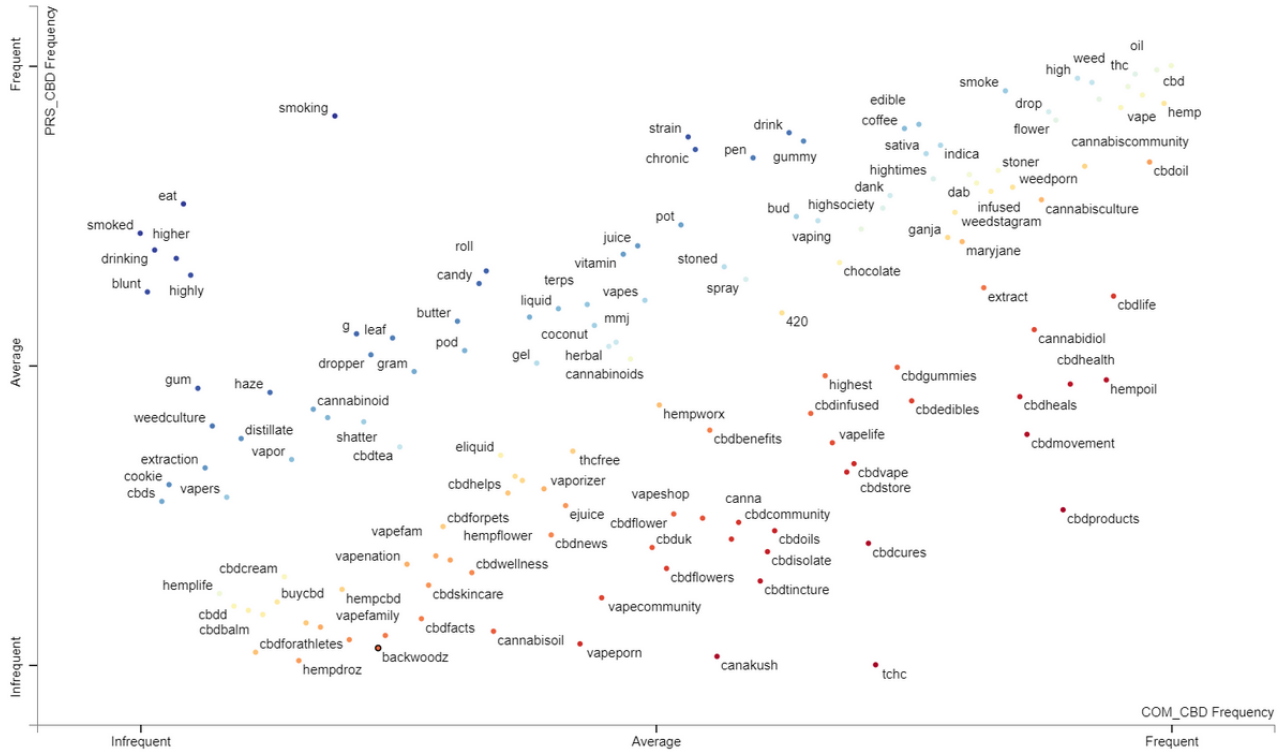
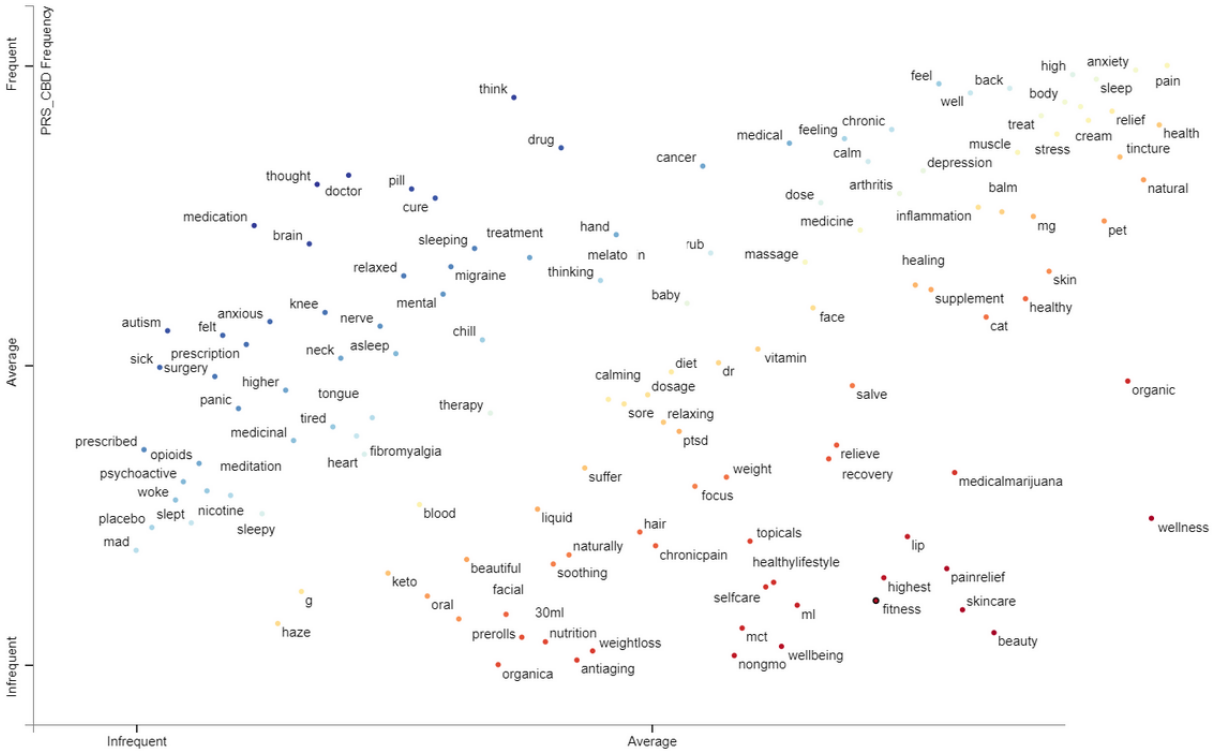


Figure 3. Other substances term frequency per class.

Figure 4. Medical/health/wellness-related term frequency per class.



Sentiment Analysis

We computed sentiment scores for both the commercial and personal CBD tweets that referenced any of the following 17 terms: *anxiety*, *anxious*, *autism*, *calm*, *calming*, *cancer*, *depression*, *energy*, *fitness*, *pain*, *pains*, *PTSD*, *skin*, *sleep*, *stress*, *weight loss*, and *wellness*. We also computed the term-level sentiment for each of these individual terms. Table 6 contains a list of the terms related to nonneutral sentiment and where the VADER score of the individual term might impact the sentiment score of the entire tweet. We calculated the sentiment for each personal and commercial CBD tweet referencing any of the 17 terms of interest, both in the original tweet text and with the term of interest removed. Using a *t* test to gauge any statistically significant difference in the mean sentient scores allowed us to

assess how sentiments about the condition itself might affect the sentiment score.

Table 6. Medical-related terms with nonneutral sentiment.

Term	VADER[a] compound score
anxiety	–0.1779
anxious	0.25
calm	0.3182
calming	0.4019
cancer	–0.6597
depression	–0.5719
energy	0.2732
pain	–0.5106
pains	–0.4215
stress	–0.4215

[a]VADER: Valence Aware Dictionary for Sentiment Reasoning.

Tables 7-9 demonstrate a significant difference in the mean sentiment score in the personal CBD tweets with the term of interest versus that obtained without the term of interest for 11 of the 17 terms examined. There was a significant difference in the mean sentiment score in the commercial CBD tweets with the term of interest versus that obtained without the term of interest for 12 of the 17 terms examined. There was also a significant difference in the mean sentiment score in the commercial CBD tweets compared to that of the commercial CBD tweets with the term of interest for 11 of the 17 terms examined, both with and without the term of interest included. Table 8 indicates that although the sentiment was overall positive, in the instances where there was a significant difference in the sentiment scores between the personal and commercial CBD tweets, the mean sentiment score of the commercial CBD scores was higher than that of the personal CBD scores. Figure 5 and Figure 6 provide examples of how the distribution of the sentiment score changed when the term of interest (“pain”) was removed from the tweet.

Table 7. Personal and commercial cannabidiol (CBD) sentiment categorical counts.

Term	Personal tweets							Commercial tweets						
	n	With term			Without term			n	With term			Without term		
		pos[a]	neu[b]	neg[c]	pos	neu	neg		pos	neu	neg	pos	neu	neg
anxiety	5353	2818	126	2409	3125	519	1718	2924	1564	44	1316	1726	352	846
anxious	515	266	11	238	307	47	161	114	53	4	57	84	5	25
autism	395	180	47	168	180	47	168	27	17	2	8	17	2	8
calm	1224	1007	17	200	761	145	318	725	659	9	57	535	80	110
calming	445	399	4	42	324	33	88	389	369	2	18	308	47	34
cancer	986	230	19	737	530	111	345	246	76	0	170	122	44	80
depression	568	164	17	387	307	31	230	326	69	8	249	178	19	129
energy	507	416	9	82	334	57	116	444	421	7	16	357	52	35
fitness	57	48	2	7	37	4	16	128	125	0	3	100	15	13
pain	7432	2948	188	4296	4985	558	1889	6287	3262	113	2912	4956	591	740
pains	394	157	11	226	225	19	150	311	168	9	134	219	11	81
ptsd[d]	217	111	14	92	111	14	92	55	33	9	13	33	9	13
skin	618	461	55	102	464	54	100	2516	2211	150	155	2229	136	151
sleep	3761	2518	356	887	2517	356	888	2980	2129	322	529	2131	321	528
stress	1012	560	18	434	713	45	254	1407	883	28	496	1100	36	271
weight loss	8	5	2	1	5	2	1	24	18	3	3	18	3	3
wellness	144	129	2	13	98	18	28	4216	4020	38	158	3106	814	296

[a]pos: positive sentiment.

[b]neu: neutral sentiment.

[c]neg: negative sentiment.

[d]ptsd: posttraumatic stress disorder.

Table 8. Personal and commercial cannabidiol (CBD) sentiment score descriptive statistics (with and without the term).

Term	Personal tweets			Commercial tweets		
	n	With term, mean (SD)	Without term, mean (SD)	n	With term, mean (SD)	Without term, mean (SD)
anxiety	5353	0.074 (0.573)	0.186 (0.557)	2924	0.118 (0.568)	0.241 (0.538)
anxious	515	0.048 (0.591)	0.203 (0.566)	114	0.08 0 (0.566)	0.254 (0.529)
autism	395	–0.001 (0.546)	–0.001 (0.546)	27	0.188 (0.557)	0.188 (0.557)
calm	1224	0.448 (0.484)	0.258 (0.540)	725	0.616 (0.374)	0.452 (0.467)
calming	445	0.608 (0.408)	0.410 (0.508)	389	0.695 (0.334)	0.513 (0.444)
cancer	986	–0.369 (0.571)	0.158 (0.559)	246	–0.303 (0.638)	0.167 (0.564)
depression	568	–0.275 (0.573)	0.122 (0.571)	326	–0.353 (0.506)	0.111 (0.525)
energy	507	0.469 (0.492)	0.324 (0.541)	444	0.681 (0.331)	0.547 (0.421)
fitness	57	0.429 (0.463)	0.263 (0.514)	128	0.633 (0.283)	0.464 (0.400)
pain	7432	–0.099 (0.605)	0.293 (0.547)	6287	0.088 (0.580)	0.490 (0.440)
pains	394	–0.098 (0.610)	0.169 (0.577)	311	0.087 (0.615)	0.342 (0.553)
ptsd[a]	217	0.037 (0.626)	0.037 (0.627)	55	0.200 (0.563)	0.200 (0.563)
skin	618	0.420 (0.501)	0.427 (0.501)	2516	0.550 (0.371)	0.568 (0.371)
sleep	3761	0.305 (0.522)	0.305 (0.523)	2980	0.392 (0.493)	0.394 (0.493)
stress	1012	0.116 (0.632)	0.360 (0.567)	1407	0.234 (0.596)	0.481 (0.396)
weight loss	8	0.289 (0.344)	0.289 (0.344)	24	0.436 (0.549)	0.436 (0.549)
wellness	144	0.606 (0.431)	0.384 (0.524)	4216	0.720 (0.279)	0.505 (0.426)

[a]ptsd: posttraumatic stress disorder.

Table 9. Personal and commercial cannabidiol (CBD) sentiment score t test results (with and without term).

Term	Personal with vs without term			Commercial with vs without term			Commercial vs personal with term			Commercial vs personal without term		
	t	*df*	*P* value	*t*	*df*	*P* value	*t*	*df*	*P* value	*t*	*df*	*P* value
anxiety	–10.31	10,704	<.001	–8.51	5846	<.001	–3.33	8275	.001	–4.28	8275	<.001
anxious	–4.29	1028	<.001	–2.39	226	.02	0.53	627	.59	–0.88	627	.38
autism	0.00	788	>.99	0.00	52	>.99	–1.74	420	.08	–1.74	420	.08
calm	9.15	2446	<.001	7.40	1448	<.001	–8.06	1947	<.001	–8.04	1947	<.001
calming	6.40	888	<.001	6.49	776	<.001	–3.37	832	.001	–3.09	832	.002
cancer	–20.71	1970	<.001	–8.65	490	<.001	–1.59	1230	.11	–0.22	1230	.83
depression	–11.67	1134	<.001	–11.49	650	<.001	2.06	892	.40	0.29	892	.77
energy	4.45	1012	<.001	5.25	886	<.001	–7.68	949	<.001	–7.02	949	<.001
fitness	1.81	112	.07	3.92	254	<.001	–3.69	183	<.001	–2.88	183	.004
pain	–41.39	14,862	<.001	–43.82	12,572	<.001	–18.37	13,717	<.001	–23.02	13,717	<.001
pains	–6.32	786	<.001	–5.43	620	<.001	–3.98	703	<.001	–4.01	703	<.001
ptsd[a]	–0.01	432	.99	0.00	108	>.99	–1.77	270	.08	–1.76	270	.08
skin	–0.24	1234	.81	–1.76	5030	.08	–7.25	3132	<.001	–7.89	3132	<.001
sleep	0.01	7520	.99	–0.12	5958	.90	–6.97	6739	<.001	–7.10	6739	<.001
stress	–9.12	2022	<.001	–11.98	2812	<.001	–4.65	2417	<.001	–5.60	2417	<.001
weight loss	0.00	14	>.99	0.00	46	>.99	–0.71	30	.48	–0.71	30	.48
wellness	3.94	286	<.001	27.38	8430	<.001	–4.72	4358	<.001	–3.35	4358	.001

[a]ptsd: posttraumatic stress disorder.

Figure 5. Distribution of sentiment scores of personal tweets referencing the term “pain.”.

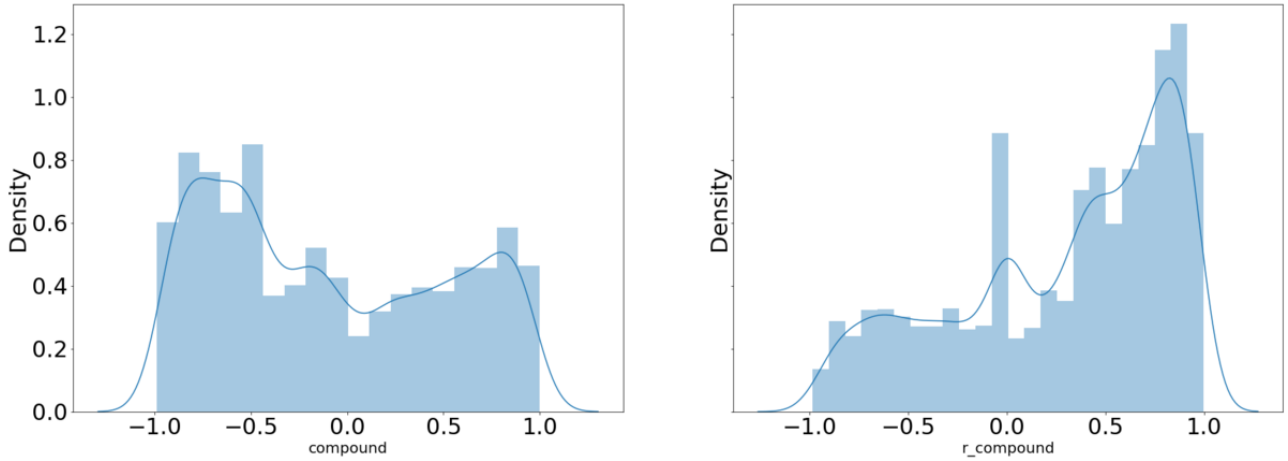
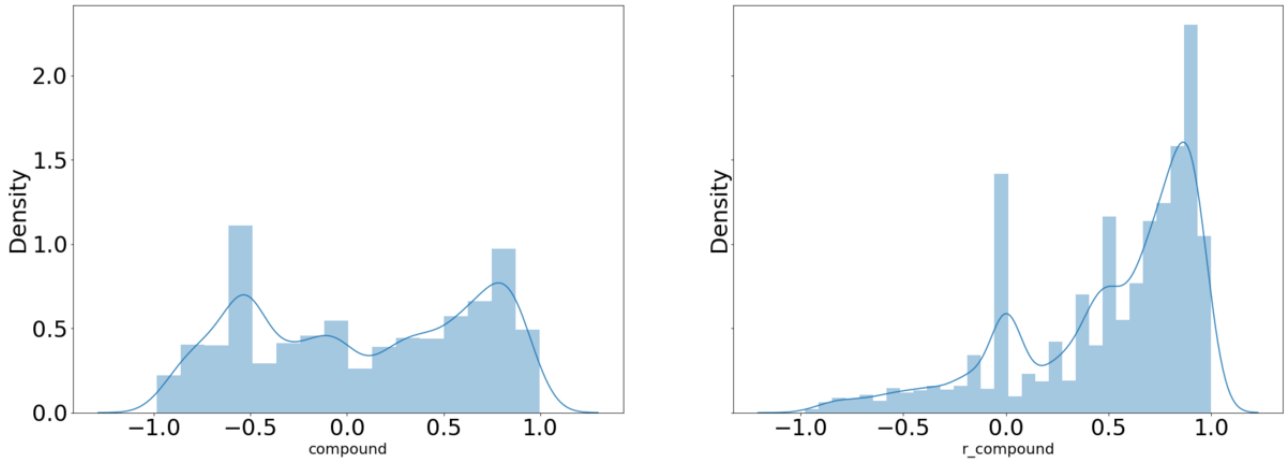
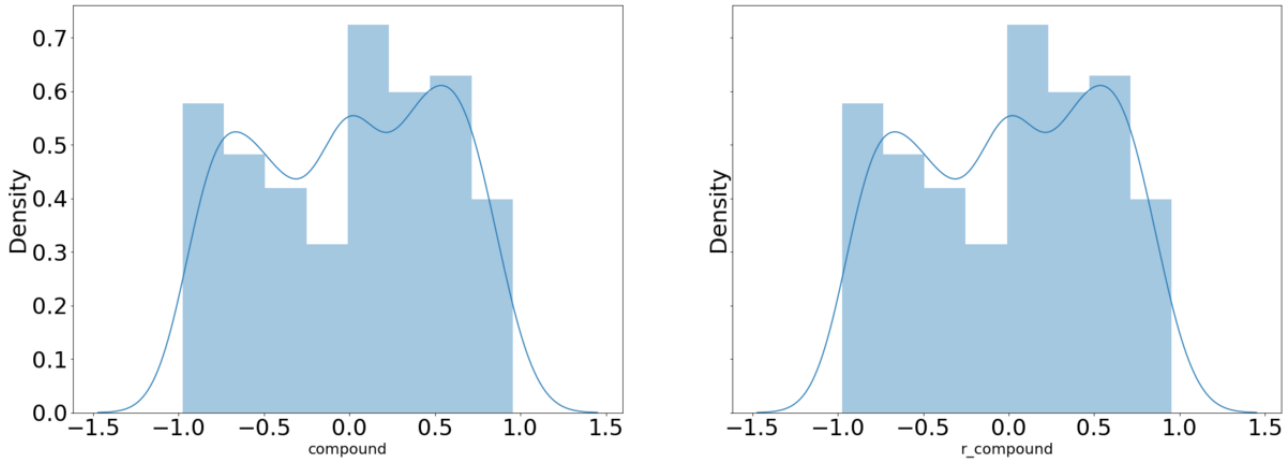


Figure 6. Distribution of sentiment scores of commercial tweets referencing the term “pain.”.



There were instances of CBD tweets that offered a mix of both positive and negative sentiments within the personal tweets, such as in tweets referencing CBD’s relationship with autism. Figure 7 shows a more negative sentiment in the personal CBD tweets referencing autism. However, the sentiment of the personal tweets did not change when the term “autism” was removed. Despite being negative, the mean sentiment score of these tweets was –0.042, which is considered neutral by the authors of the VADER model. We observed a large amount of personal CBD tweets referencing the term “autism,” with 42.5% (168/395) of personal CBD tweets referencing autism classified as negative versus 45.6% (180/395) being classified as positive and 11.9% (47/395) classified as neutral.

Figure 7. Sentiment distribution of personal cannabidiol-related tweets referencing autism.



Textbox 3 shows tweets that our classifiers identified to be CBD-related that contained the word “autism.” These personal CBD tweets sometimes favored and sometimes disfavored CBD as a treatment for autism. In the commercial CBD tweets referencing autism, we observed both implicit and explicit claims regarding CBD’s ability to treat autism.

Our framework thus works well in contexts where the efficacy claims of medications and supplements are both validated and refuted.

Textbox 3. Examples of personal and commercial tweets referencing cannabidiol (CBD) and autism (paraphrased slightly for anonymity).

Personal autism CBD tweets

@user @user He's...on a thc/cbd tincture. It's helped his autism beautifully

@user I use CBD for my C-PTSD [posttraumatic stress disorder] and the overstimulation that comes from Autism...it's kinda hard to function...It works better than any antipsychotic I've ever been on.

I see a lot of ppl on my timeline...claiming CBD can "cure" autism is bad and so is anyone knowingly peddling the idea what is wrong with you people URL

@user @user @user if you work somewhere that lies about curing autism with cbd oil you should feel bad

Commercial autism CBD tweets

10 Best CBD Oils For Autism - URL 10-bestcbd-oils-for-autism-40/

CBD INFUSED … BOTTLES @ FRUIT PASTELS, THESE ARE IDEAL FOR KIDS WITH ADHD, AUTISM ETC #THECBDCHEMIST #GOODNIGHTSLEE

COMPANY_NAME® Announces Autism Hope Alliance Sponsorship URL #cannabis #hemp #cbd #vape #cbdoil #natural #anxiety #pain #stress #health #pharma #wellness #beauty #domains URL

People use CBD to treat everything from epilepsy and autism to chronic pain and anxiety. URL

Discussion

Principal Findings

Text classification of tweets provides a means to segment tweets into defined groups at a large scale. We have demonstrated that we can do this with tweets related to CBD by using text classification to identify tweets that reflect personal usage of CBD and tweets that reflect the sales and/or commercialization of CBD. This classification of public social media data is useful because CBD has not been subjected to the same tests and clinical trials as modern medications, yet is currently being used to treat a variety of conditions without proof of safety or efficacy. Our analysis provided a methodology to identify the terms of interest that are frequently referenced in the commercial and personal corpora of CBD tweets, as well as a comparison of these term frequencies in relation to the document class (commercial or personal CBD). This allowed us to identify the medical conditions commonly referenced in both document classes at high frequencies, as well as terms that occur more frequently in one document class over the other. We also used the VADER model to analyze the sentiment of personal CBD tweets referencing certain medical conditions and symptoms. Despite the US FDA's warnings regarding the marketing and promotion of CBD as treatment of autism and Alzheimer disease, while certainly not the most frequent conditions referenced, we did observe multiple instances of these tactics.

These methods and results speak to the recent efforts of researchers to use social networks to analyze the concept of misinformation in ways that are directly related to the potential problem of CBD misinformation. Ferrand et al [25] analyzed responses to queries from common digital assistants such as Siri, Alexa, and Google for misinformation regarding vaccines. Chen et al [26] collected social network posts from Weibo related to cancer and observed that 30% of posts contained misinformation. Ahmed et al [27] collected tweets referencing COVID-19 and 5G, and performed graph analysis to identify and analyze how misinformation was being disseminated online. Allem et al [3] observed unsubstantiated health claims related to cannabis on Twitter. More recently, Rovetta and Bhagavathula [28] also observed an abundance of COVID-19 misinformation in their analysis of tweets.

To address potential social media misinformation in the area of CBD, it is important to develop methods for gathering and classifying text corpora. Previous studies using the internet and social media have described the personal and commercial discourses about CBD. Narayanan et al [15] made use of internet-based data sources to examine CBD trends by examining Google searches, demonstrating that interest in CBD oil increased significantly from 2014 to 2018. Tran and Kavuluru [31] used CBD-related posts from Reddit and comments submitted to the FDA regarding these posts to examine the conditions that are commonly being treated by CBD. The researchers in this study examined both corpora of texts for medical conditions and methods of use in posts and comments using the term "CBD," along with any indication of therapy implied in the two corpora.

There have also been nonmachine learning approaches to researching marijuana sentiment on Twitter, such as the work done by Nguyen et al [32]. Their study collected marijuana-related tweets, disregarded the tweets that were authored by less influential posters, manually annotated the marijuana tweets on a Likert scale via crowdsourcing, and segmented them by demographics applied to the data set through a proprietary service. The researchers observed more promarijuana attitudes among African Americans and youth/younger adults. In another example of a research approach that did not rely primarily on machine learning, Krauss et al [33] based their marijuana sentiment analysis on crowdsourced tweets. The researchers aimed to examine the preferences between marijuana and alcohol on Twitter. They collected tweets containing alcohol and marijuana references, and then annotated the tweets via crowdsourcing. The results showed that 54% of

the tweets normalized marijuana and alcohol, 24% showed a preference for marijuana over alcohol, 2% showed a preference for alcohol over marijuana, 7% showed negative sentiment toward both alcohol and marijuana, and 13% demonstrated no sentiment toward either substance.

Our proposed framework extends the existing CBD research by further examining the perceptions of CBD in online discussions through a comparison of the terms and sentiment of tweets that reflect the personal use of CBD and tweets that reflect the sales and/or promotion of CBD. No other studies have attempted this type of comparative work, and this approach helped to examine which terms are being used proportionally or disproportionally, and to compare the sentiment of personal and commercial CBD tweets. Our methods can be applied to other types of research aimed at analyzing trends in the consumption and advertising of unregulated substances.

Conclusion and Future Work

The strengths of our framework are the ability to identify personal and commercial CBD tweets, associated conditions, and sentiment. However, some limitations should be noted. First, we limited our search to tweets referencing "CBD" and "cannabidiol." Our preliminary research did not indicate that the topic of CBD is as subjected to slang terms as other forms of cannabis (eg, THC) and did not indicate that it was necessary to include additional related terms in the search [34]. Additionally, we limited our collection to tweets as Twitter is one of the world's largest social networks that provides the ability to collect a large volume of data quickly. Second, our data were collected over an approximate 3-month period. Although the data collection period was relatively small, we were able to identify trends in personal and commercial CBD tweets that will be useful for future studies. Another limitation is the use of the dictionaries (standard, slang, and SNOMED-CT) for finding associated medical conditions. This step was based on checking high-frequency terms against dictionaries to determine medical relevancy. Future research could use deep neural network models to extract medical-related named entities from the tweets to automate and to possibly obtain context in which the medically related term is being used [35]. Finally, although we did not explicitly identify and remove social bots from our collection, as discussed by Himelein-Wachowiak et al [36], we did remove bots during the annotation process, as tweets that were possibly machine-generated were not considered personal or commercial CBD tweets.

We successfully used text classification to identify tweets making personal or commercial CBD references. When we applied two classifiers to the collection of tweets, we identified multiple medical conditions, body parts, symptoms, other substances, and cannabis references that were mentioned at high frequencies in both the personal and commercial CBD corpora, as well as conditions that were mentioned disproportionately in one corpora over the other. This suggests that CBD is being used and marketed for consistent types of ailments. Our sentiment analysis showed that the term of interest can indeed influence the sentiment score; when controlling for the term, 15 of 17 terms tested showed a positive sentiment within the personal CBD tweets and all 17 terms showed a positive sentiment within the commercial CBD tweets. This suggests that CBD, on the whole, is well-regarded in terms of its medical applications and that the commercial claims are not gross distortions of popular sentiment; however, we observed evidence where claims may have been exaggerated. We encourage future research to investigate the patterns in sentiment, usage, and sales of CBD as well as other forms of cannabis over time. Additionally, we recommend extending this proposed framework by further using text mining and machine learning methods to identify the dissemination of misinformation as it relates to CBD's health and medical benefits.

Conflicts of Interest

None declared.

References

1. Aggarwal SK, Carter GT, Sullivan MD, ZumBrunnen C, Morrill R, Mayer JD. Medicinal use of cannabis in the United States: historical perspectives, current trends, and future directions. J Opioid Manag 2009 Jan 29;5(3):153-168. [doi: 10.5055/jom.2009.0016] [Medline: 19662925]
2. Campos AC, Moreira FA, Gomes FV, Del Bel EA, Guimarães FS. Multiple mechanisms involved in the large-spectrum therapeutic potential of cannabidiol in psychiatric disorders. Philos Trans R Soc Lond B Biol Sci 2012 Dec 05;367(1607):3364-3378 [FREE Full text] [doi: 10.1098/rstb.2011.0389] [Medline: 23108553]
3. Allem J, Escobedo P, Dharmapuri L. Cannabis surveillance with Twitter data: emerging topics and social bots. Am J Public Health 2020 Mar;110(3):357-362. [doi: 10.2105/ajph.2019.305461]
4. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem J. Twitter posts about cannabis use during pregnancy and postpartum: a content analysis. Subst Use Misuse 2021;56(7):1074-1077. [doi: 10.1080/10826084.2021.1906277] [Medline: 33821757]
5. FDA and cannabis: Research and Drug Approval Process. US Food and Drug Administration. 2020 Oct. URL: https://www.fda.gov/news-events/public-health-focus/fda-and-cannabis-research-and-drug-approval-process [accessed 2020-12-02]
6. Palmieri B, Laurino B, Vadalà M. A therapeutic effect of cbd-enriched ointment in inflammatory skin diseases and cutaneous scars. Clin Ter 2019;170(2):e93-e99 [FREE Full text] [doi: 10.7417/CT.2019.2116] [Medline: 30993303]
7. Linares IMP, Guimaraes FS, Eckeli A, Crippa ACS, Zuardi AW, Souza JDS, et al. No acute effects of cannabidiol on the sleep-wake cycle of healthy subjects: a randomized, double-blind, placebo-controlled, crossover study. Front Pharmacol 2018 Apr 5;9:315. [doi: 10.3389/fphar.2018.00315] [Medline: 29674967]

8. Masataka N. Anxiolytic effects of repeated cannabidiol treatment in teenagers with social anxiety disorders. Front Psychol 2019 Nov 8;10:2466. [doi: 10.3389/fpsyg.2019.02466] [Medline: 31787910]
9. Shannon S, Lewis N, Lee H, Hughes S. Cannabidiol in anxiety and sleep: a large case series. Perm J 2019;23:18-041 [FREE Full text] [doi: 10.7812/TPP/18-041] [Medline: 30624194]
10. Skelley JW, Deas CM, Curren Z, Ennis J. Use of cannabidiol in anxiety and anxiety-related disorders. J Am Pharm Assoc (2003) 2020;60(1):253-261. [doi: 10.1016/j.japh.2019.11.008] [Medline: 31866386]
11. Boyaji S, Merkow J, Elman RNM, Kaye AD, Yong RJ, Urman RD. The role of cannabidiol (CBD) in chronic pain management: an assessment of current evidence. Curr Pain Headache Rep 2020 Jan 24;24(2):4. [doi: 10.1007/s11916-020-0835-4] [Medline: 31980957]
12. Johnson JR, Burnell-Nugent M, Lossignol D, Ganae-Motan ED, Potts R, Fallon MT. Multicenter, double-blind, randomized, placebo-controlled, parallel-group study of the efficacy, safety, and tolerability of THC:CBD extract and THC extract in patients with intractable cancer-related pain. J Pain Symptom Manage 2010 Feb;39(2):167-179 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.06.008] [Medline: 19896326]
13. Überall MA. A review of scientific evidence for THC: CBD oromucosal spray (Nabiximols) in the management of chronic pain. J Pain Res 2020;13:399-410. [doi: 10.2147/JPR.S240011] [Medline: 32104061]
14. Kosgodage US, Mould R, Henley AB, Nunn AV, Guy GW, Thomas EL, et al. Cannabidiol (CBD) is a novel inhibitor for exosome and microvesicle (EMV) release in cancer. Front Pharmacol 2018 Aug 13;9:889. [doi: 10.3389/fphar.2018.00889] [Medline: 30150937]
15. Narayanan S, Lazar Neto F, Tanco K, Lopez G, Liu W, Bruera E, et al. Cannabidiol (CBD) oil, cancer, and symptom management: a Google trends analysis of public interest. J Altern Complement Med 2020 Apr 01;26(4):346-348 [FREE Full text] [doi: 10.1089/acm.2019.0428] [Medline: 31971821]
16. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H. Recognizing depression from Twitter activity. 2015 Presented at: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18, 2015; Seoul, Korea p. 3187-3196. [doi: 10.1145/2702123.2702280]
17. Carchiolo V, Longheu A, Malgeri M. Using Twitter data and sentiment analysis to study diseases dynamics. 2015 Presented at: ITBAM 2015: Information Technology in Bio- and Medical Informatics; September 3-4, 2015; Valencia, Spain. [doi: 10.1007/978-3-319-22741-2_2]
18. Oscar N, Fox PA, Croucher R, Wernick R, Keune J, Hooker K. Machine learning, sentiment analysis, and tweets: an examination of Alzheimer's disease stigma on Twitter. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):742-751. [doi: 10.1093/geronb/gbx014] [Medline: 28329835]
19. Abualigah L, Alfar HE, Shehab M, Hussein AMA. Sentiment analysis in healthcare: a brief review. In: Abd Elaziz M, Al-qaness M, Ewees A, Dahou A, editors. Recent Advances in NLP: The Case of Arabic Language. Studies in Computational Intelligence, vol 874. Cham: Springer; 2020:129-141. [doi: 10.1007/978-3-030-34614-0_7]
20. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. 2015 Presented at: 8th International Conference on Weblogs and Social Media (AAAI); June 2014; Ann Arbor, MI.
21. Kirlić A, Orhan Z. Measuring human and Vader performance on sentiment analysis. Invent J Res Technol Engineer Manag 2017;1(12-4):42-46.
22. Elbagir S, Yang J. Twitter sentiment analysis using Natural Language Toolkit and VADER sentiment. 2019 Presented at: International MultiConference of Engineers and Computer Scientists (IMECS 2019); March 13-15, 2019; Hong Kong. [doi: 10.1142/9789811215094_0005]
23. Newman H, Joyner D. Sentiment analysis of student evaluations of teaching. In: Penstein RC, editor. Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science. Volume 10948. Cham: Springer; 2018:246-250. [doi: 10.1007/978-3-319-93846-2_45]
24. Park C, Seo D. Sentiment analysis of Twitter corpus related to artificial intelligence assistants. 2018 Presented at: 2018 5th International Conference on Industrial Engineering and Applications (ICIEA); April 26-28, 2018; Singapore. [doi: 10.1109/iea.2018.8387151]
25. Ferrand J, Hockensmith R, Houghton RF, Walsh-Buhi ER. Evaluating smart assistant responses for accuracy and misinformation regarding human papillomavirus vaccination: content analysis study. J Med Internet Res 2020 Aug 03;22(8):e19018 [FREE Full text] [doi: 10.2196/19018] [Medline: 32744508]
26. Chen L, Wang X, Peng TQ. Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]
27. 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]
28. 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]
29. Kuehn BM. FDA, FTC warn companies claiming their supplements treat infertility. JAMA 2021 Jul 06;326(1):20. [doi: 10.1001/jama.2021.10551] [Medline: 34228081]
30. Kessler J. Scattertext: a browser-based tool for visualizing how corpora differ. 2017 Presented at: Association for Computational Linguistics 2017, System Demonstrations; 2017; Vancouver, Canada. [doi: 10.18653/v1/p17-4015]

31. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. Int J Drug Policy 2020 Mar;77:102688 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102688] [Medline: 32092666]
32. Nguyen A, Hoang Q, Nguyen H, Nguyen D, Tran T. Evaluating marijuana-related tweets on Twitter. 2017 Presented at: 7th Annual Computing and Communication Workshop and Conference; January 9-11, 2017; Las Vegas, NV. [doi: 10.1109/ccwc.2017.7868364]
33. Krauss MJ, Grucza RA, Bierut LJ, Cavazos-Rehg PA. "Get drunk. Smoke weed. Have fun.": A content analysis of tweets about marijuana and alcohol. Am J Health Promot 2016 Nov 17;31(3):200-208. [doi: 10.4278/ajhp.150205-qual-708]
34. Turner J, Kantardzic M. Twitter query expansion via Word2Vec-Urban Dictionary Model. 2018 Presented at: International Conference on Computing and Big Data; 2018; Charleston, SC. [doi: 10.1145/3277104.3278310]
35. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. 2019. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-12-06]
36. Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz HA, et al. Bots and misinformation spread on social media: implications for COVID-19. J Med Internet Res 2021 May 20;23(5):e26933 [FREE Full text] [doi: 10.2196/26933] [Medline: 33882014]

Abbreviations

CBC: cannabichromene
CBD: cannabidiol
FDA: Food and Drug Administration
MCT: medium-chain triglyceride
PTSD: posttraumatic stress disorder
SNOMED-CT: Systemized Nomenclature of Medicine-Clinical Terms
TF-IDF: term frequency-inverse document frequency
THC: tetrahydrocannabinol
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by R Kukafka, G Eysenbach; submitted 20.01.21; peer-reviewed by J Lyu, JP Allem; comments to author 10.04.21; revised version received 21.05.21; accepted 10.11.21; published 20.12.21.

Please cite as:
Turner J, Kantardzic M, Vickers-Smith R
Infodemiological Examination of Personal and Commercial Tweets About Cannabidiol: Term and Sentiment Analysis
J Med Internet Res 2021;23(12):e27307
URL: https://www.jmir.org/2021/12/e27307
doi:10.2196/27307
PMID:34932014

Original Paper

Characterizing and Identifying the Prevalence of Web-Based Misinformation Relating to Medication for Opioid Use Disorder: Machine Learning Approach

Mai ElSherief[1], PhD; Steven A Sumner[2], MD; Christopher M Jones[3], DPH; Royal K Law[4], PhD; Akadia Kacha-Ochana[2], MPH; Lyna Shieber[5], DPhil; LeShaundra Cordier[5], MPH; Kelly Holton[3], MPH; Munmun De Choudhury[6], PhD

[1]University of California, San Diego, San Diego, CA, United States

[2]Office of Strategy and Innovation, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, United States

[3]National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, United States

[4]Division of Injury Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, United States

[5]Brunet-García, Atlanta, GA, United States

[6]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Munmun De Choudhury, PhD
School of Interactive Computing
Georgia Institute of Technology
756 W Peachtree St NW
Atlanta, GA, 30308
United States
Phone: 1 4043858603
Email: munmund@gatech.edu

Abstract

Background: Expanding access to and use of medication for opioid use disorder (MOUD) is a key component of overdose prevention. An important barrier to the uptake of MOUD is exposure to inaccurate and potentially harmful health misinformation on social media or web-based forums where individuals commonly seek information. There is a significant need to devise computational techniques to describe the prevalence of web-based health misinformation related to MOUD to facilitate mitigation efforts.

Objective: By adopting a multidisciplinary, mixed methods strategy, this paper aims to present machine learning and natural language analysis approaches to identify the characteristics and prevalence of web-based misinformation related to MOUD to inform future prevention, treatment, and response efforts.

Methods: The team harnessed public social media posts and comments in the English language from Twitter (6,365,245 posts), YouTube (99,386 posts), Reddit (13,483,419 posts), and Drugs-Forum (5549 posts). Leveraging public health expert annotations on a sample of 2400 of these social media posts that were found to be semantically most similar to a variety of prevailing opioid use disorder–related myths based on representational learning, the team developed a supervised machine learning classifier. This classifier identified whether a post's language promoted one of the leading myths challenging addiction treatment: that the use of agonist therapy for MOUD is simply replacing one drug with another. Platform-level prevalence was calculated thereafter by machine labeling all unannotated posts with the classifier and noting the proportion of myth-indicative posts over all posts.

Results: Our results demonstrate promise in identifying social media postings that center on treatment myths about opioid use disorder with an accuracy of 91% and an area under the curve of 0.9, including how these discussions vary across platforms in terms of prevalence and linguistic characteristics, with the lowest prevalence on web-based health communities such as Reddit and Drugs-Forum and the highest on Twitter. Specifically, the prevalence of the stated MOUD myth ranged from 0.4% on web-based health communities to 0.9% on Twitter.

Conclusions: This work provides one of the first large-scale assessments of a key MOUD-related myth across multiple social media platforms and highlights the feasibility and importance of ongoing assessment of health misinformation related to addiction treatment.

(J Med Internet Res 2021;23(12):e30753) doi:10.2196/30753

KEYWORDS

opioid use disorder; substance use; addiction treatment; misinformation; social media; machine learning; natural language processing

Introduction

Background

In the United States, opioid overdose continues to be a leading cause of death [1]. The Centers for Disease Control and Prevention estimates that the total economic burden of prescription opioid misuse in the country alone is US $78.5 billion a year, including the costs of health care, lost productivity, treatment, and criminal justice involvement [2]. Alarmingly, opioid overdoses increased by 30% from July 2016 to September 2017 in 52 areas in 45 US states [3]. Consequently, in 2017, the Department of Health and Human Services declared it as a public health emergency [4]. Central to addressing the opioid crisis is expanding access to medication treatment for opioid use disorder (MOUD) [5]. MOUD increases treatment retention and reduces opioid use, risk behaviors that transmit blood-borne pathogens, and overdose mortality [6]. However, despite its well-documented effectiveness, studies have found that MOUD is underused due in part to stigma and misperceptions about treatment [7].

In recent years, many individuals have been seeking both conventional and nonconventional ways to recover from substance use, including using web-based resources [8]. For these conditions, as well as opioid use disorder (OUD), research has shown that individuals turn to the web for promoting and discovering recovery strategies, for example, appropriating the Forum77 forum for prescription drug use recovery [9] and participating in 12-step programs such as Narcotics Anonymous [10,11]. Social support is another motivation behind individuals with substance use disorders turning to social media; Rubya and Yarosh [12] examined peer support for substance use disorder recovery meetings through video chat, discovering that video chat support groups not only provide immediacy and convenience in meeting needs but can also be places of obtaining emotional and informational support. More recently, researchers have examined patterns of anonymity in web-based recovery communities [13]. Specific to OUD, previous studies have investigated the different types of web-based discourse associated with opioid use, including personal use, whether it is associated with legitimate use or abuse of opioids [14], or whether it involves the promotion of clinically unverified treatments [15]. Abuse discourse on social media platforms has been further broken down into stand-alone use and co-use of opioids with other opioids, illicit drugs, and alcohol [16]. In addition, a prior study analyzed the web-based discourse surrounding the perception of opioids [17]. The perception of opioids included commentary on the opioid crisis, opioids in general, and interaction with news surrounding the opioid crisis or medical use of opioids [17]. Researchers in the past have also harnessed social media data as unobtrusive sensors to identify individuals who might benefit from or be receptive to treatment and recovery interventions [18]. Others have computationally examined and compared web-based discussion communities to discover the intent to contribute to web-based mental health communities [19]. In general, social media platforms have been found to allow increased self-disclosure for users to discuss otherwise sensitive and stigmatizing topics such as OUD [20]. Apart from self-disclosure, social media data provide unique opportunities for understanding the users' sentiments and opinions [21], which may be insightful from the perspective of addiction treatment.

Despite the positive benefits of social media, existing attempts of individuals with OUD are often challenged because of the pervasiveness of inaccurate and potentially harmful health misinformation on social media platforms [15]. Health misinformation is defined as a health-related claim of a fact that is currently false because of a lack of scientific evidence [22]. In general, misinformation is usually attributed to misconceptions and is not intended to cause harm. Disinformation is false information that is created deliberately to cause harm, with motivations that are often social, political, or financial. Although misinformation and disinformation are inherently false, malinformation is usually based on real information that is taken completely out of or without context to inflict harm [23]. Fake news is defined as fabricated information that mimics news media content in form but not in organizational process or intent [24,25]. Molina et al [24] have outlined key indicators of fake news such as content that is not fact-checked, is emotionally charged, is written in narrative style, has unverified sources, or comes from an unknown source. In this study, we focused on the language of false claims surrounding MOUDs regardless of intent; therefore, it might be the case that we captured a few instances of disinformation, possibly on web-based platforms that lack constant domain-specific moderation. Thus, we use the term *health misinformation* as we assume that the spread of these claims is not intentional.

From the discourse on infectious disease outbreaks and global epidemics to alternative therapies to tackle behavioral health problems, web-based misinformation can have adverse effects on public health, including negatively influencing people's health literacy, attitudes, beliefs, and health-related decision-making [22]. For example, antivaccine-promoting social media posts legitimize debate about vaccine safety, contribute to reductions in vaccination rates, and increase vaccine-preventable diseases such as measles [26]. In the context of public health crises, social media rumors circulating during the Ebola outbreak in 2014 were found to create hostility toward health workers, which posed challenges in controlling the epidemic [27]. Most recently, the novel COVID-19 pandemic has come to be defined by a tsunami of persistent misinformation to the public on everything from the utility of masks and the effectiveness of social distancing to even the promise of vaccines, together contributing to an increased

COVID-19 pandemic burden [28]. At-risk populations are known to be particularly vulnerable to misinformation [22,29] because of a lack of reliable information outside of formal clinical or rehabilitation contexts [30,31]. In fact, studies show that because of exposure to such misinformation, people worry that they will be ostracized by their community if their substance use is revealed to others, thus delaying treatment [32].

Given the limited uptake of MOUD, the potential contribution of health misinformation to this public health problem, and the fact that information about barriers to MOUD is challenging to ascertain from other data sources, exploring digital health-seeking behavior through *passive sensing* of misinformation related to MOUD provides an important avenue for addressing this problem. Thus, *infodemiology*, which refers to the science of studying the distribution and determinants of information and user-generated content in an electronic medium such as the web in general and social media in particular [33], has the opportunity to shape MOUD-related health promotion strategies and policies. Given the potential impact of misinformation in the midst of the ongoing overdose crisis, there is a critical need to better understand misinformation-related social media posts on OUD treatment. In fact, in recent years, approaches in infodemiology have been noted to be important in mitigating public health problems stemming from *infodemics* [34,35], a portmanteau of *information* and *epidemic* that typically refers to a rapid and far-reaching spread of both accurate and inaccurate information about a disease.

Objective

In this study, we focus on one particular myth (and its language variants) related to MOUD: *agonist therapy or medication-assisted treatment (MAT) is simply replacing one drug with another*. For example, someone might express this myth by saying "You are not really in recovery if you are on Suboxone." This myth is believed to be one of the major reasons cited for individual hesitancy to initiate MOUD; it has been discussed extensively in clinical literature [29,36,37] and has been discredited by evidence that MOUDs facilitate recovery and that multiple other chronic health conditions such as diabetes and asthma necessitate reliance on daily medication to maintain health.

By adopting a multidisciplinary, mixed methods strategy, this paper aims to present the first work that investigates the characteristics and prevalence of web-based misinformation related to MOUD across 3 types of web-based social platforms to inform future prevention, treatment, and response efforts. Our contributions include a set of machine learning (ML) models that classify whether a post revolves around conversations surrounding a specific MOUD as replacing one drug with another or explorations of lexical variations characterizing web-based conversations relating to this myth.

Methods

Data Set Curation

We first identified and curated a set of clinically grounded and publicly prevalent myths that surround OUD treatment and developed a lexicon of opioid-related keywords associated with different aspects of OUD. We captured different types of opioids, such as natural opiates, semisynthetic opioids, and synthetic opioids, and included opioids that were over-the-counter, prescription based, or illicit. For each generic name, we also included trade and combination product names in consultation with the substance use literature and the public health coauthors. This resulted in a total of 152 keywords curated in the lexicon. We then curated a diverse data set from Twitter, YouTube, and the web-based health communities Reddit and Drugs-Forum. These platforms were selected as (1) they are adopted pervasively by Americans and (2) there are well-established means and infrastructures for collecting meaningful data sets by leveraging app programming interfaces to query them and access public posts on these platforms. According to the Pew Research Center, in 2021, 18% of US adults use Reddit, 23% use Twitter, and 81% use YouTube [38]. In addition, these platforms have been mined in prior substance abuse literature for abuse monitoring and digital epidemiology purposes [39-41]. For all the platforms we investigated, we focused on public posts and messages created between January 1, 2018, and December 31, 2019.

Our data set collection methodology for Twitter comprised querying for all tweets that included 1 of the words in our lexicon. This process yielded a total of 6,365,245 tweets. For YouTube, owing to limitations in the number of comments that can be accessed, we restricted the 152 keywords to 11 OUD treatment keywords such as *buprenorphine* and *naltrexone*. We used the YouTube app programming interface to identify 552 YouTube videos that contained 1 of the 11 keywords in the title and then collected all of the associated comments (99,386 comments). We relied on expert domain knowledge to identify subforums pertinent to OUD for Reddit and Drugs-Forum and used the full set of 152 keywords for these sites. For Reddit, we used data from 22 opioid-specific subreddits: *r/Carfentanil*, *r/opiates*, *r/fentanyl*, *r/opiatesmemorial*, *r/modquittingkratom*, *r/Methadone*, *r/suboxone*, *r/kratom*, *r/heroin*, *r/quittingkratom*, *r/Tianeptine*, *r/loperamide*, *r/naltrexone*, *r/oxycodone*, *r/OpiatesRecovery*, *r/Opiatewithdrawal*, *r/lean*, *r/heroinaddiction*, *r/HeroinHeroines*, *r/OpiateChurch*, *r/suboxone*, and *r/OurOverUsedVeins*. This resulted in a total of 1,189,590 posts and 12,293,829 comments. In addition, we collected all 5549 messages posted under the *Opiates and Opioids* subforums on Drugs-Forum [42]. Throughout the paper, we have combined Reddit and Drugs-Forum content under the category of web-based health communities, as both have similar structure, format, and affordances.

ML Approach Using Expert Involvement

Web-based discourse surrounding OUD is semantically rich; that is, there are different words and combinations of words that people use to convey meaning. Previous literature has quantitatively and qualitatively investigated various categories of language pertaining to OUD, including OUD use (own use, use by others, abuse, legitimate use, and co-use), OUD perception (commentary on opioid crisis or opioids in general), and OUD advertisements [14,16,17]. In light of such linguistic richness and prior investigations, we adopted an ML and natural

language analysis methodology to identify posts relevant to the myth under investigation in the huge search space.

We first leveraged representation learning techniques, which are a set of techniques that allow a system to automatically discover the representations needed for feature detection or classification from raw data [43] to construct document-level embeddings (consisting of 4096 dimensions) of the myth statement noted earlier. For this, we used a bidirectional long short-term memory (LSTM) sentence encoder model universally trained on a natural language inference task [44]. LSTM was a suitable choice here as it allowed us to learn long-term dependencies among words in sentence structures. We then used this model to encode all the collected posts. Following this step, we obtained the k-nearest neighbor (KNN), where k=200, for semantically most similar posts per platform for the MOUD-related seed myth under investigation. Second, using a mixed methods approach, our models then harnessed qualitative content analysis in the form of public health expert annotations to label a total of 800 posts (200 KNNs per platform) and annotate whether each post was relevant to the myth (ie, whether the post discussed MOUD and described MOUD as using one drug to replace another). Hence, we modeled this problem as a binary classification task where the positive class denoted a post discussing the aforementioned piece of misinformation and the negative class represented any post that was not relevant to the myth. Each myth KNN post was annotated by the same expert public health annotator to provide consistent annotations within the linguistic domain of a given myth.

Leveraging these annotations as training data, we finally built and evaluated a series of supervised ML models, ranging from logistic regression (LR) and support vector machines to feedforward neural networks and LSTM networks. Our feature set included lexical features such as n-grams (n=1, 2, 3), term frequency–inverse document frequency (TF–IDF) weights, and representation learning features, including sentence-based embeddings (semantic) and transformer-based embeddings, such as bidirectional encoder representations from transformers [45] and bidirectional encoder representations from transformers for biomedical text mining [46]. We used all annotations belonging to our myth and considered all the samples from other myths as negative training samples. On the basis of this process, we obtained 171 positive samples and 2229 negative samples. Owing to this large imbalance, we leveraged an oversampling technique from the rare class, called the synthetic minority oversampling technique [47]. We then split the data set into training and test samples with an 80% to 20% split, respectively. We leveraged 2 techniques for cross-validation: k-fold cross-validation (for LR and support vector machine models) and an independent validation sample to tune a model's hyperparameters (for the LSTM model).

Results

Table 1 and Figure 1 show the best-performing ML models in terms of their area under the curve, precision, recall, and F_1 scores. Our best-performing model was a combination of TF-IDF features and an LR classifier, achieving a precision of 0.85, a recall of 0.91, an F_1 score of 0.88, and an area under the curve of 0.9. By applying our best-performing model to machine label all posts in our data sets, we were able to estimate the prevalence of posts related to the myth under investigation on each platform. The prevalence of posts among our sampled comments that were related to the myth that the use of MOUD does not constitute true recovery was 0.4%, 0.9%, and 0.58% for web-based health communities, Twitter, and YouTube, respectively. For additional context and interpretability in terms of how our best-performing models operated per platform, 2 examples of posts that were classified correctly by our classifier are provided in Table 2, along with the top words used by the classifier to attain a relevancy decision for each post on each platform. Here we observed some consistencies in the discussions of the myth across platforms. For example, we noted that our model was able to pick up on the use of verbs synonymous with *replac*, such as *switch*, which was not originally included in the myth phrasing. In addition, the verb *go* was used in multiple contexts, such as going to Alcoholics Anonymous meetings instead of relying on MATs and going through withdrawals from MAT. We also noted the presence of multiple drug names such as *Ativan*, *buprenorphine*, *methadone*, and *suboxone*.

Table 1. Macroperformance metrics of the opioid use disorder treatment myth classifiers[a].

Model	Accuracy	AUC[b]	Precision	Recall	F_1 score
LR[c]+semantic[d]	0.84	0.84	0.87	0.86	0.86
LR+TF-IDF[e]	0.91	0.9	0.85	0.91	0.88
LSTM[f]+BERT[g]	0.77	0.81	0.83	0.84	0.84

[a]Training and test data drawn from 2400 opioid-related posts from Twitter, web-based health communities, and YouTube.

[b]AUC: area under the curve.

[c]LR: logistic regression.

[d]InferSent semantic representations (4096 features).

[e]TF-IDF: term frequency-inverse document frequency.

[f]LSTM: long short-term memory.

[g]BERT: bidirectional encoder representations from transformer.

Figure 1. Receiver operating characteristic (ROC) curves for each classifier. Training and test data drawn from 2400 opioid-related posts from Twitter, web-based health communities, and YouTube. (A) logistic regression+semantic; (B) logistic regression+term frequency-inverse document frequency; and (C) long short-term memory+bidirectional encoder representations from transformers.

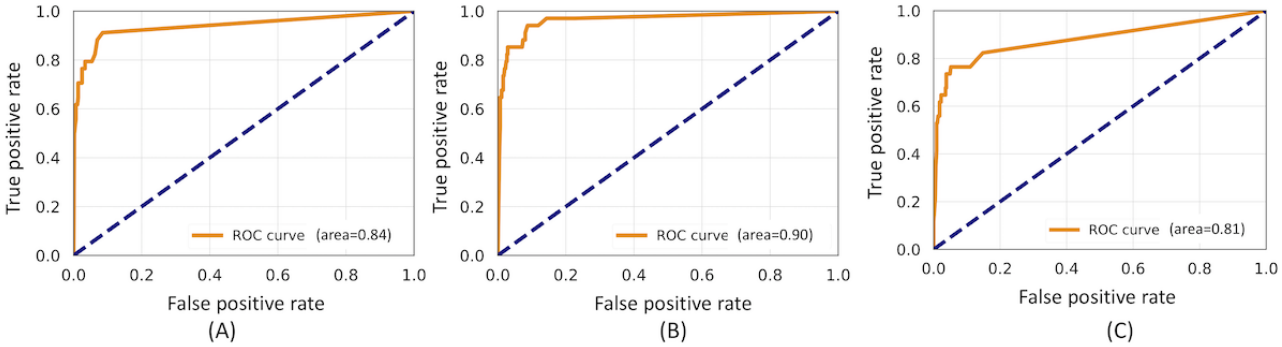


Table 2. Paraphrased examples detected by our best-performing classifier on different platforms and top features highlighted. Raw posts are paraphrased to prevent traceability and author identification.

Platform and raw paraphrased post	Preprocessed post	Feature power	
		Contribution	Feature
Web-based health communities			
"Don't take the kratom. Don't switch one drug for another. Go to an aa meeting. for real. IV Ativan is usually the go to drug for such symptoms."	"take kratom switch one drug anoth go aa meeting for real iv ativan usual go drug symptom"	• +2.955 • +2.055 • +1.783 • +1.710 • +1.585 • +1.479 • +1.251 • +1.238 • +1.057	• ativan • go • usual drug • aa anoth one • symptom switch
"[...] [Name of a person] said: Please dont take it! If you can stop using opiates and not go back just go through the withdrawals. If you would trust me, you dont want the withdrawals (especially long term) that Bupe has! Please know that the length of the withdrawal period for maintenance users is in part dependent on the dose [...]"	"[...] [Name of a person] said pleas take stop use opiat go back go withdraw promis want withdraw especi long term bupe pleas understand length withdraw period mainten user part dose depend [...]"	• +1.080 • +0.734 • +0.693 • +0.691 • +0.559 • +0.510 • +0.475 • +0.460 • +0.435 • +0.427	• therapi • buprenorphin dose • mainten replac • go • need appropri pain • may
Twitter			
"Saying that people dying of Heroin/Fentanyl ODs[a] is because they are getting Rx meds from doctor is just irresponsible & untrue. When someone gets addicted to methadone, what is happening is that $$ from the street are getting switched to $$ to big Pharma & our GOV. Abusing methadone/Suboxone still leads to deaths."	"say rx med dr mani die heroin fentanyl od simpli irrespons untru get someon addict methadon simpli switch street go big pharma gov peopl still die abus methadon suboxon"	• +2.41 • +1.750 • +1.142 • +0.814 • +0.752 • +0.732 • +0.589 • +0.398 • +0.376	• methadon • simpli med suboxon switch go irrespons mani • get big
"I wonder if the w/d from Bupe or Suboxone is any easier than heroin or fentanyl. Let's say someone switched to MAT[b] as an interim because they wanted to be substance free; do you think they would go through w/d 2x?"	"w bupe suboxon easier heroin fentanyl one want substanc free would one go w x one switch mat interim"	• +4.095 • +3.801 • +1.418 • +1.157 • +1.071 • +1.041 • +0.922 • +0.583 • +0.188	• mat • one • bupe suboxon switch go easier substanc heroin want
YouTube			
"Okay I am planning to discontinue treatment. I feel I need support, but with my family disapproving of this treatment of being on MMT[c], I don't seem to be getting that. To them, it is no different from doing heroin everyday. They say I am switching one addiction for another [...]"	"decid discontinu treatment famili agre form treatment im get support mmt dont see differ heroin everyday say switch one addict anoth [...]"	• +3.722 • +3.595 • +1.857 • +1.290 • +1.076 • +0.577 • +0.393 • +0.385 • +0.332	• methadon • treatment anoth • go suboxon form one mmt • get switch

Platform and raw paraphrased post	Preprocessed post	Feature power	
		Contribution	Feature
"Your fear of the withdrawal symptoms is totally legit. They suck. Did you tell your doctor about your intake of the prescription? There needs to be some sort of a planned approach for not just quitting, but also to make sure you ween off your meds properly. Have you heard of Suboxone? It's a prescription medication that basically will help you with withdrawals as well as give you a crutch. Kratom is another option, but going through the withdrawal alone and learning how to walk away as a substance-free person takes a lot of daring and audacity, so you need to have what it takes for it."	"fulli understand fear withdraw symptom suck doctor know intak prescript game plan set quit also effort ween med sure heard suboxon prescript medic short summari itll help withdraw well act like crutch anoth thing kratom go withdraw one one learn walk away medfre person take lot gut courag take"	• +1.370 • +1.009 • +0.878 • +0.810 • +0.780 • +0.704 • +0.678 • +0.658 • +0.626 • +0.606	• one • prescript med • anoth medic • quit symptom effort suboxon set

[a]OD: overdose.

[b]MAT: medication-assisted treatment.

[c]MMT: methadone maintenance treatment.

The top 10 features (terms) associated with our best-performing model (LR+TF-IDF) for identifying relevant posts and their TF-IDF values are shown in Table 3. These terms include *mat*, *assist*, *treatment*, *replac*, *therapi*, *rehab*, *methadon*, *behavior*, *habit*, and *substitut*. Furthermore, to provide additional insight into words used by the ML model to identify myth-related posts, for each of the top 10 terms, we display the 15 words with the closest semantic proximity (based on training a Word2Vec embedding model [48]) as measured by cosine similarity. Qualitative assessment of the identified words revealed excellent identification of synonymous terms and phrases, including those that were unlikely to be readily suggested or identified by human readers, such as *ost* (opioid substitution therapy).

Table 3. Top 10 salient features and their associated Word2Vec model nearest neighbors per platform[a].

Feature and platform	Nearest neighbors
mat (14.9)	
Web-based health communities	assist (0.49), proven (0.46), lifer (0.46), abstin (0.42), recoveri (0.41), stigma (0.41), mmt (0.41), superior (0.4), vivitrol (0.39), align (0.39), treatment (0.39), lifesav (0.39), mainten (0.39), adhes (0.39), bamboo (0.38)
Twitter	treatment (0.61), medic (0.48), suboxon (0.46), bupe (0.43), need (0.4), therapi (0.39), behavior (0.39), stigmat (0.38), oud (0.38), postod (0.38), clear (0.37), suffici (0.37), med (0.36), part (0.36)
YouTube	assist (0.72), recommend (0.72), care (0.7), bullshit (0.68), recoveri (0.67), truli (0.66), mention (0.66), step (0.65), anyon (0.65), mani (0.64), could (0.64), oud (0.63), possibl (0.63), lose (0.62), integr (0.62)
assist (12.44)	
Web-based health communities	mat (0.49), counsel (0.45), profession (0.44), supervis (0.42), lifesav (0.4), help (0.38), certifi (0.38), vivitrol (0.38), aftercar (0.37), florida (0.37), longterm (0.37), mainten (0.37), recoveri (0.36), consult (0.36), transit (0.36)
Twitter	appropri (0.4), profession (0.37), switzerland (0.36), mat (0.36), aaap (0.35), grade (0.34), staff (0.34), necessary (0.33), ongo (0.33), treatment (0.33), discrimin (0.32), center (0.31), evidenc (0.31)
YouTube	famili (0.79), medic (0.77), judg (0.73), mat (0.72), recommend (0.71), lose (0.71), mani (0.71), could (0.7), therapi (0.7), battl (0.69), wonder (0.69), truli (0.67), win (0.67), recoveri (0.65), group (0.63)
treatment (11.43)	
Web-based health communities	program (0.52), evid (0.51), ibogain (0.51), nation (0.51), medic (0.5), assess (0.49), longterm (0.48), wherein (0.48), addict (0.48), establish (0.47), intervent (0.46), protocol (0.46), rehabilit (0.46), observ (0.46), augment (0.46)
Twitter	medic (0.67), therapi (0.66), mat (0.61), use (0.6), postod (0.59), need (0.55), drug (0.53), opioid (0.52), methadone (0.52), patient (0.51), reduc (0.48), rehab (0.48), provid (0.46), prescrib (0.45)
YouTube	individu (0.72), treat (0.65), truli (0.65), ibogain (0.64), acknowledg (0.64), oud (0.62), recoveri (0.62), comfort (0.62), assist (0.61), receiv (0.6), great (0.6), keep (0.59), wonder (0.59), bullshit (0.56), worri (0.55)
replac (9.91)	
Web-based health communities	swap (0.44), substitut (0.41), exercis (0.39), switch (0.39), fix (0.38), hormon (0.38), lifestyl (0.37), atom (0.37), still (0.36), healthi (0.35), discomfort (0.35), slowli (0.34), bad (0.34), lead (0.33), use (0.33)
Twitter	substitut (0.48), altern (0.42), simpli (0.37), adjunct (0.35), extrem (0.35), swap (0.35), scienc (0.34), type (0.34), neither (0.32), panacea (0.32), creat (0.32), reduc (0.32), result (0.32), lifelong (0.31), grade (0.3)
YouTube	hip (0.74), due (0.58), lot (0.57), altern (0.55), complet (0.55), result (0.54), k (0.54), rapid (0.53), someth (0.51), realiti (0.49), exchang (0.49), would (0.48), anti (0.47), argu (0.47), told (0.47)
therapi (9.43)	
Web-based health communities	counsel (0.61), cbt (0.57), trauma (0.54), dbt (0.54), somat (0.49), therapist (0.48), ptsd (0.47), aftercar (0.46), tool (0.46), cognit (0.46), treatment (0.46), adjunct (0.45), psychiatri (0.45), longterm (0.45)
Twitter	treatment (0.66), medic (0.44), psycholog (0.43), sizabl (0.43), psychosoci (0.43), acupunctur (0.43), use (0.42), postod (0.41), howev (0.41), incl (0.4), mat (0.39), success (0.37), pain (0.37), odb (0.37), need (0.37)
YouTube	group (0.81), recoveri (0.76), na (0.74), requir (0.71), assist (0.7), oud (0.69), famili (0.69), recommend (0.67), aa (0.66), set (0.66), individu (0.64), base (0.64), great (0.63), bullshit (0.59), mat (0.59)
rehab (8.45)	
Web-based health communities	inpati (0.59), facil (0.55), detox (0.55), centr (0.51), outpati (0.51), relaps (0.49), iop (0.49), ua (0.49), sober (0.48), homeless (0.47), residenti (0.47), jail (0.46), program (0.46), na (0.46), voluntarili (0.44)
Twitter	treatment (0.48), residenti (0.46), mandatori (0.43), get (0.39), staffer (0.38), drug (0.38), go (0.37), one (0.37), clean (0.35), sobrieti (0.34), need (0.33), whitewash (0.33), mostli (0.33), let (0.33)
YouTube	went (0.83), show (0.77), gone (0.76), new (0.7), bottom (0.67), bare (0.67), littl (0.63), day (0.62), gonna (0.62), sadli (0.6), away (0.6), process (0.59), gave (0.59), mom (0.54), keep (0.53)
methadon (8.43)	
Web-based health communities	suboxon (0.78), heroin (0.57), opiat (0.57), bupe (0.52), oxi (0.5), clinic (0.5), sub (0.49), taper (0.49), mainten (0.48), mmt (0.48), dope (0.45), stigma (0.45), detox (0.45), addict (0.45)

Feature and platform	Nearest neighbors
Twitter	treatment (0.52), opioid (0.46), drug (0.46), medic (0.42), use (0.41), postod (0.39), base (0.38), residenti (0.38), option (0.37), continu (0.37), provid (0.37), mani (0.36), client (0.36)
YouTube	trust (0.68), switch (0.66), without (0.65), scare (0.62), suboxon (0.61), im (0.6), hate (0.59), due (0.59), anyway (0.56), year (0.56), dose (0.53), transit (0.53), wait (0.51), yr (0.51), center (0.51)
behavior (8.14)	
Web-based health communities	behaviour (0.52), empathi (0.49), eif (0.44), repetit (0.44), undetect (0.44), destruct (0.44), hostil (0.43), cbt (0.43), exhibit (0.43), pattern (0.42), drugseek (0.42), flexibl (0.42), manipul (0.42)
Twitter	physic (0.39), behaviour (0.39), mat (0.38), topamax (0.37), workflow (0.36), yoga (0.36), cognit (0.35), nprzyb (0.35), multilevel (0.35), recogn (0.35), rank (0.34), diseas (0.33), group (0.33), kneepain (0.33), approach (0.33)
YouTube	jail (0.9), interest (0.89), servic (0.87), grant (0.87), integr (0.84), organ (0.83), learn (0.8), via (0.79), find (0.77), healthcar (0.77), health (0.77), final (0.75), set (0.74), mani (0.72), educ (0.71)
habit (7.96)	
Web-based health communities	struggl (0.52), willpow (0.5), allen (0.48), carr (0.48), smoke (0.48), stop (0.45), habit (0.45), cig (0.45), cigarette (0.44), feel (0.42), go (0.42), definit (0.41), sobrieti (0.4), time (0.4), smoker (0.4)
Twitter	crack (0.39), googlawaqpp (0.37), dailyrecord (0.36), pushi (0.36), rehab (0.33), bright (0.33), intox (0.33), black-watch (0.32), mccain (0.32), filthi (0.32), iff (0.31), weed (0.31), sober (0.31)
YouTube	herion (0.74), beer (0.58), slave (0.54), codein (0.54), trade (0.53), chemic (0.52), far (0.52), issu (0.52), kratom (0.51), compound (0.5), anoth (0.49), wake (0.49), immedi (0.49), sick (0.48), evil (0.48)
substitut (7.65)	
Web-based health communities	deriv (0.47), replac (0.45), sert (0.45), synthes (0.44), indol (0.44), halogen (0.43), amin (0.43), keton (0.41), phenyl (0.41), monocycl (0.41), hydrogen (0.4), piperidin (0.4), haloalkyl (0.39)
Twitter	replac (0.47), ost (0.35), psilocybin (0.33), dcr (0.31), lesser (0.31), hepatitisc (0.3), licat (0.3), abstain (0.29), deaden (0.29), halflif (0.28), assist (0.28), cab (0.28)
YouTube	anoth (0.69), sell (0.69), address (0.67), none (0.67), slave (0.65), exchang (0.63), isnt (0.61), what (0.61), crutch (0.6), issu (0.59), sinc (0.58), there (0.58), trade (0.57), meant (0.55), unbroken (0.54)

[a]Data from 112,281 opioid-related posts identified by our best-performing model from Twitter, web-based health communities, and YouTube. The first column depicts the features and their term frequency-inverse document frequency scores. The nearest neighbors column also depicts the cosine similarity between each word and the corresponding feature. Words in posts are stemmed before being fed to models (eg, recovery is stemmed to its root *recoveri*). Web-based health communities refer to Reddit and Drugs-Forum.

Discussion

Principal Findings

Harms propagated by misinformation are aplenty on the web and come at both financial and societal costs. People often accept what they read as true, especially if it comes from a reasonably reputable source, and do not question the information, no matter how astounding or alarming. In fact, people even repeat the more remarkable information regardless of how accurate it is. In the context of MOUD, it can lead to grave consequences, including overdose deaths [29]. To the best of our knowledge, this is the first study to examine MOUD-related misinformation on a large scale, harnessing conversations happening on the web.

Closely related to our work is the study by Jamison et al [49], which leverages a collection of tweets to quantify vaccine misinformation. Similar to our work, Jamison et al [49] coded tweets into thematic categories based on vaccine sentiment (positive, negative, or neutral). However, our work leveraged thematic categories (relevant and not relevant to the myth) to design ML-based models that are able to identify misinformation in the context of MOUDs. Heimer et al [29] discussed prevalent misconceptions about OUDs in the United States through 3 crises (1865-1913, 1960-1975, and 1995-today). Similar to our focus, the authors acknowledged opioid *abstinence-based* recovery models as a prevailing misconception and promoted the large-scale expansion of MAT. Our work complements their work by investigating this misconception quantitatively through the lens of social media. Chenworth et al [50] investigated the perception of the general public toward methadone and buprenorphine-naloxone on Twitter. The authors identified that a common barrier to treatment with these medications was the idea of opioid substitution—the exchange of one opioid addiction for another [50]. Our work investigates this barrier at a deeper level by building models that are able to recognize this type of discourse on social media.

Our results have important public health implications. Across multiple platforms, we detected that the prevalence of posts about a single myth related to medication treatment for OUD in our sample ranged from 4 per 1000 posts on web-based health communities to 9 per 1000 posts on Twitter. This is notable, as, at any time, there are likely multiple myths being discussed on the web, suggesting that the total volume of misinformation content related to opioids may be a substantial proportion of the total posts. The prevalence of such information has not been

previously quantified, and this study offers important insights into the potential scope of this health information issue.

Although we cannot speculate on the exact reason why Twitter presented more misinformation in the case of OUD-related misinformation as that requires causal inference analysis, which is beyond the scope of this paper, prior literature has pointed out the lack of active expert or clinical-based moderation on Twitter [51]. Although web-based health communities are also not immune to bad behavior and antisocial activities such as trolling, spamming, and harassment, these communities are often guided by strict norms against such behavior and moderated to ensure the quality and credibility of the content being shared [52]. Prior studies on different types of web-based health communities have demonstrated that adequate active moderation increases the engagement of members and consequently also increases the beneficial outcomes for members in a web-based community [53]. In fact, the moderators themselves regard their moderation style as important for the regulation and stimulation of membership engagement [54,55]. We suspect that, because of these established moderation norms, we observed a relatively lesser prevalence of MOUD misinformation in the web-based communities we studied. We noted that Twitter does implement some broad governance rules that allow for certain types of information to stay on the platform, whereas others are removed (eg, graphic violence and adult content [56]). The platform also has provisions to tackle the widespread presence of hate speech and abusive content [57]. However, to the best of our knowledge, Twitter does not implement policies toward the moderation of MOUD misinformation. Our conjecture is that, because of this existing practice, our study revealed a greater prevalence of this misinformation on the platform. Nevertheless, in light of the ongoing COVID-19 pandemic, Twitter has broadened its definition of harm to address "content that goes directly against guidance from authoritative sources of global and local public health information" [58]. We hope that the findings of this study can motivate social media platforms to consider moderation approaches toward substance misuse information as well.

Given the significant prevalence of myths around OUD treatment, as shown in this study, a possible approach to counter web-based misinformation could be to perform targeted, expert fact-checking of social media posts. This could mirror and harness guidelines adopted by public health organizations to debunk unverified information about OUD treatment. For instance, substance use experts can be identified and asked to review the content of social media posts to determine their accuracy. These experts could critically appraise a post and produce a response comprising a lay summary of the evidence in addition to a detailed, referenced evidence review. This review could be directly linked to the original post through appropriate platform affordances to provide users with quick access to fact-checked information. Specific fact-checking processes could also be tailored to individual social media platforms, given the differences we observed both in terms of prevalence and the linguistic characteristics of the myth discussions. Qualitative exploration of the characteristics of the statements identified by the ML approach revealed linguistic and topical diversity. Some statements explicitly referenced the main concept we queried for—that MOUD represents replacing one drug with another. However, related statements were identified in which alternative treatments such as kratom entered into the discussion. Rationales for hesitancy toward MOUD also became apparent, including concerns about the addictiveness of MOUD, the nature of withdrawal symptoms from MOUD, and concerns about industry or governmental motivations for recommending MOUD. Understanding these concerns is directly relevant to providing health information, understanding the role of digital information ecosystems as a supplant or adjuvant resource in substance misuse treatment, and addressing treatment hesitancy.

In addition to fact-checking efforts, public health engagement campaigns could also be used to address specific cases of misinformation. Recent research suggests that information campaigns led by trusted community members and health partners can help address health misinformation on social platforms [59]. Accordingly, alliances can be forged with social media influencers and key opinion leaders to run targeted health promotion campaigns. Interventions such as those with positive messaging can also be tailored to the preferences, perceptions, and cultures of different platforms. Educational interventions that improve literacy around OUD treatment and reduce the stigma that precludes seeking help, as well as ecologically sensitive interventions that open up avenues to access social support, could also empower individuals to be better equipped to deal with OUD treatment myths on the web. In short, although the literature on strategies to effectively counter health misinformation is still emerging, at minimum, this work highlights the importance of ongoing assessment and awareness of what health information is being prominently discussed on the web to guide both the provision of effective health care and public health prevention activities.

We note some limitations of this work. Although our analysis included large data sets from diverse web-based platforms, MOUD-related discussions happen on a wide variety of social platforms, and the prevalence of misinformation across a broader set of web-based environments needs characterization. For one platform, YouTube, limitations in the number of comments that can be accessed required restriction of the keyword list, which may have affected the prevalence of misinformation, although the estimate from YouTube was comparable with the other platforms. Furthermore, this research did not examine the nature of conversations surrounding the OUD treatment myth we focused on in this paper, such as whether a conversation might be reinforcing or countering the myth or discussing other previously known myths. Future work may unpack these characteristics of web-based discussions while also investigating additional myths about OUD misuse that surface on web-based platforms. Finally, geospatial-temporal studies on MOUD misinformation that originates and spreads via social media platforms can be a promising and significant direction for future research; they can influence interventions such as targeted location-based misinformation-countering campaigns as well as help clinicians respond to patients' false beliefs or misperceptions.

Conclusions

Using ML and natural language analysis, our research demonstrated promise in identifying social media posts that centered on treatment myths about OUD, including how these discussions varied across platforms in terms of prevalence. As the overdose epidemic continues to evolve, attention from health professionals to health information on the web that drives patient decision-making will continue to be a critical element of prevention.

Acknowledgments

This research was supported through a contract by the Centers for Disease Control and Prevention (CDC; via the Department of Health and Human Services) to the Georgia Institute of Technology (principal investigator: MDC). CDC investigators participated in the design and conduction of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. ME and MDC had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. ME conducted this research while being affiliated to the Georgia Institute of Technology. The authors thank Kenneth Kannampully for his help with the Drugs-Forum.com data collection. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Conflicts of Interest

None declared.

References

1. Florence C, Luo F, Rice K. The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. Drug Alcohol Depend 2021 Jan 01;218:108350. [doi: 10.1016/j.drugalcdep.2020.108350] [Medline: 33121867]
2. Florence CS, Zhou C, Luo FM, Xu L. The economic burden of prescription opioid overdose, abuse, and dependence in the United States, 2013. Med Care 2016 Oct;54(10):901-906 [FREE Full text] [doi: 10.1097/MLR.0000000000000625] [Medline: 27623005]
3. Vivolo-Kantor AM, Seth P, Gladden RM, Mattson CL, Baldwin GT, Kite-Powell A, et al. Vital Signs: trends in Emergency Department visits for suspected opioid overdoses - United States, July 2016-September 2017. MMWR Morb Mortal Wkly Rep 2018 Mar 09;67(9):279-285 [FREE Full text] [doi: 10.15585/mmwr.mm6709e1] [Medline: 29518069]
4. Alexander LM, Keahey D, Dixon K. Opioid use disorder: a public health emergency. JAAPA 2018 Oct;31(10):47-52. [doi: 10.1097/01.JAA.0000545072.09344.ee] [Medline: 30252764]
5. Strategy to combat opioid abuse, misuse, and overdose : a framework based on the five point strategy. U.S. Department of Health and Human Services. 2018. URL: https://www.hsdl.org/?abstract&did=816001 [accessed 2021-03-15]
6. Blanco C, Volkow ND. Management of opioid use disorder in the USA: present status and future directions. Lancet 2019 Apr;393(10182):1760-1772. [doi: 10.1016/s0140-6736(18)33078-2]
7. Olsen Y, Sharfstein JM. Confronting the stigma of opioid use disorder--and its treatment. J Am Med Assoc 2014 Apr 09;311(14):1393-1394. [doi: 10.1001/jama.2014.2147] [Medline: 24577059]
8. Seelye KQ. In heroin crisis, white families seek gentler war on drugs. The New York Times. 2015. URL: https://www.nytimes.com/2015/10/31/us/heroin-war-on-drugs-parents.html [accessed 2021-12-09]
9. MacLean D, Gupta S, Lembke A, Manning C, Heer J. Forum77: An analysis of an online health forum dedicated to addiction recovery. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.: ACM; 2015 Jan 01 Presented at: 18th ACM Conference on Computer Supported Cooperative Work & Social Computing; March 14, 2015; Vancouver, Canada p. 1511-1526. [doi: 10.1145/2675133.2675146]
10. Rubya S. Facilitating peer support for recovery from substance use disorders. In: Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2017 Jan 01 Presented at: CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 2017; Denver, Colorado p. 172-177. [doi: 10.1145/3027063.3048431]
11. Campbell SW, Kelley MJ. Mobile phone use among alcoholics anonymous members: new sites for recovery. New Media Society 2008 Dec 01;10(6):915-933. [doi: 10.1177/1461444808096251]
12. Rubya S, Yarosh S. Video-mediated peer support in an online community for recovery from substance use disorders. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Jan 01 Presented at: ACM Conference on Computer Supported Cooperative Work and Social Computing; February 2017; Denver, Colorado p. 1454-1469. [doi: 10.1145/2998181.2998246]
13. Rubya S, Yarosh S. Interpretations of online anonymity in alcoholics anonymous and narcotics anonymous. Proc ACM Hum-Comput Interact 2017 Dec 06;1(CSCW):1-22. [doi: 10.1145/3134726]
14. Hu H, Phan N, Geller J, Vo H, Manasi B, Huang X, et al. Deep self-taught learning for detecting drug abuse risk behavior in tweets. In: Proceedings of the International Conference on Computational Social Networks. 2018 Jul 13 Presented at: International Conference on Computational Social Networks; December 18-20, 2018; Shanghai, China p. 330-342. [doi: 10.1007/978-3-030-04648-4_28]

15. Chancellor S, Nitzburg G, Hu A, Zampieri F, De Choudhury M. Discovering alternative treatments for opioid use recovery using social media. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Apr 15 Presented at: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK p. 1-15. [doi: 10.1145/3290605.3300354]
16. Lossio-Ventura JA, Bian J. An inside look at the opioid crisis over Twitter. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2018 Dec 03 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec. 3-6, 2018; Madrid, Spain p. 1496-1499. [doi: 10.1109/bibm.2018.8621101]
17. Tibebu S, Chang VC, Drouin C, Thompson W, Do MT. At-a-glance - What can social media tell us about the opioid crisis in Canada? Health Promot Chronic Dis Prev Can 2018 Jun;38(6):263-267 [FREE Full text] [doi: 10.24095/hpcdp.38.6.08] [Medline: 29911824]
18. Eshleman R, Jha D, Singh R. Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM).: IEEE; 2017 Jan 01 Presented at: IEEE international conference on bioinformatics and biomedicine (BIBM); November 13-16, 2017; Kansas City, MO p. 849-854. [doi: 10.1109/bibm.2017.8217766]
19. 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]
20. Young SD, Koussa M, Lee S, Perez H, Gill N, Gelberg L, et al. Feasibility of a social media/online community support group intervention among chronic pain patients on opioid therapy. J Addict Dis 2018 Jan 05;37(1-2):96-101 [FREE Full text] [doi: 10.1080/10550887.2018.1557992] [Medline: 30614403]
21. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Lee KE, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. Drug Alcohol Depend 2017 Dec 01;181:5-10 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]
22. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. J Am Med Assoc 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
23. Wardle C, Derakhshan H. Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. 2017 Jan 15. URL: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html [accessed 2021-12-09]
24. Molina MD, Sundar SS, Le T, Lee D. “Fake News” is not simply false information: a concept explication and taxonomy of online content. Am Behav Scient 2019 Oct 14;65(2):180-212. [doi: 10.1177/0002764219878224]
25. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
26. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/ajph.2018.304567]
27. Jones B, Elbagir N. Are myths making the ebola outbreak worse. CNN World. 2014. URL: https://edition.cnn.com/2014/08/20/world/africa/ebola-myths/index.html [accessed 2021-12-09]
28. Scientific American. URL: https://www.scientificamerican.com/article/covid-misinformation-is-killing-people1/ [accessed 2021-02-15]
29. Heimer R, Hawk K, Vermund SH. Prevalent misconceptions about opioid use disorders in the united states produce failed policy and public health responses. Clin Infect Dis 2019 Jul 18;69(3):546-551 [FREE Full text] [doi: 10.1093/cid/ciy977] [Medline: 30452633]
30. Wu L, Blazer DG, Li T, Woody GE. Treatment use and barriers among adolescents with prescription opioid use disorders. Addict Behav 2011 Dec;36(12):1233-1239 [FREE Full text] [doi: 10.1016/j.addbeh.2011.07.033] [Medline: 21880431]
31. Schnoll SH. Misconceptions and realities of the prescription opioid epidemic. Clin Pharmacol Ther 2018 Jun 21;103(6):963-965. [doi: 10.1002/cpt.1033] [Medline: 29464692]
32. Woo J, Bhalerao A, Bawor M, Bhatt M, Dennis B, Mouravska N, et al. "Don't Judge a Book Its Cover": a qualitative study of methadone patients' experiences of stigma. Subst Abuse 2017 Mar 23;11:1178221816685087 [FREE Full text] [doi: 10.1177/1178221816685087] [Medline: 28469424]
33. 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]
34. Rothkopf DJ. SARS also spurs an 'information epidemic'. Newsday. 2003. URL: https://www.newsday.com/sars-also-spurs-an-information-epidemic-1.269835 [accessed 2021-12-09]
35. Eysenbach G. Infodemiology: the epidemiology of (mis)information. Am J Med 2002 Dec;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0]
36. Wakeman SE, Barnett ML. Primary care and the opioid-overdose crisis — buprenorphine myths and realities. N Engl J Med 2018 Jul 05;379(1):1-4. [doi: 10.1056/nejmp1802741]

37. Grinspoon P. 5 myths about using Suboxone to treat opiate addiction. Harvard Health Publishing. 2021. URL: https://www.health.harvard.edu/blog/5-myths-about-using-suboxone-to-treat-opiate-addiction-2018032014496 [accessed 2021-03-15]
38. Social media fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/?menuItem=3814afe3-3f3c-4623-910b-8a6a37885ab8 [accessed 2021-01-15]
39. Sarker A, DeRoos AB, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. J Am Med Inform Assoc 2020 Feb 01;27(2):315-329 [FREE Full text] [doi: 10.1093/jamia/ocz162] [Medline: 31584645]
40. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1-5 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
41. Sarker A, Gonzalez-Hernandez G, DeRoos F, Perrone J. Towards real-time opioid abuse surveillance: machine learning for automatic characterization of opioid-related tweets. In: Proceedings of the 39th International Congress of the European Association of Poisons Centres and Clinical Toxicologists (EAPCCT). 2019 Jan 01 Presented at: 39th International Congress of the European Association of Poisons Centres and Clinical Toxicologists (EAPCCT); May 21-24, 2019; Naples, Italy p. 475. [doi: 10.1080/15563650.2019.1598646]
42. Drugs-Forum. URL: https://drugs-forum.com/forums/opiates-opioids.33/ [accessed 2021-02-15]
43. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A. Supervised learning of universal sentence representations from natural language inference data. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen, Denmark p. 670-680. [doi: 10.18653/v1/d17-1070]
44. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
45. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-12-09]
46. 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 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
47. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
48. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013 Jan 01. URL: https://arxiv.org/abs/1301.3781 [accessed 2021-12-09]
49. Jamison A, Broniatowski DA, Smith MC, Parikh KS, Malik A, Dredze M, et al. Adapting and extending a typology to identify vaccine misinformation on Twitter. Am J Public Health 2020 Oct;110(S3):331-339. [doi: 10.2105/ajph.2020.305940]
50. Chenworth M, Perrone J, Love JS, Graves R, Hogg-Bremer W, Sarker A. Methadone and suboxone mentions on Twitter: thematic and sentiment analysis. Clin Toxicol (Phila) 2021 Nov 06;59(11):982-991. [doi: 10.1080/15563650.2021.1893742] [Medline: 33821724]
51. Peters ME, Uible E, Chisolm MS. A Twitter education: why psychiatrists should tweet. Curr Psychiatry Rep 2015 Dec 13;17(12):94. [doi: 10.1007/s11920-015-0635-4] [Medline: 26463050]
52. Saha K, Ernala S, Dutta S, Sharma E, De Choudhury M. Understanding moderation in online mental health communities. In: Proceedings of the International Conference on Human-Computer Interaction. 2020 Jan 01 Presented at: International Conference on Human-Computer Interaction; July 19-24, 2020; Copenhagen, Denmark p. 87-107. [doi: 10.1007/978-3-030-49576-3_7]
53. Matzat U, Rooks G. Styles of moderation in online health and support communities: an experimental comparison of their acceptance and effectiveness. Comput Hum Behav 2014 Jul;36:65-75. [doi: 10.1016/j.chb.2014.03.043]
54. Coulson NS, Shaw RL. Nurturing health-related online support groups: exploring the experiences of patient moderators. Comput Hum Behav 2013 Jul;29(4):1695-1701. [doi: 10.1016/j.chb.2013.02.003]
55. Skousen T, Safadi H, Young C, Karahanna E, Safadi S, Chebib F. Successful moderation in online patient communities: inductive case study. J Med Internet Res 2020 Mar 17;22(3):e15983 [FREE Full text] [doi: 10.2196/15983] [Medline: 32181743]
56. Twitter Rules. URL: https://help.twitter.com/en/rules-and-policies/twitter-rules [accessed 2021-02-15]
57. Offensive Tweets and Content. URL: https://help.twitter.com/en/safety-and-security/offensive-tweets-and-content [accessed 2021-02-15]
58. Baker SA, Wade M, Walsh MJ. The challenges of responding to misinformation during a pandemic: content moderation and the limitations of the concept of harm. Media Int Australia 2020 Aug 17;177(1):103-107. [doi: 10.1177/1329878x20951301]
59. Bonnevie E, Rosenberg SD, Kummeth C, Goldbarg J, Wartella E, Smyser J. Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine. PLoS One 2020 Oct 16;15(10):e0240828 [FREE Full text] [doi: 10.1371/journal.pone.0240828] [Medline: 33064738]

Abbreviations

CDC: Centers for Disease Control and Prevention
KNN: k-nearest neighbor
LR: logistic regression
LSTM: long short-term memory
MAT: medication-assisted treatment
ML: machine learning
MOUD: medication for opioid use disorder
OUD: opioid use disorder
TF-IDF: term frequency-inverse document frequency

Edited by M Focsa; submitted 27.05.21; peer-reviewed by M Lotto, X Cheng; comments to author 10.07.21; revised version received 04.10.21; accepted 19.10.21; published 22.12.21.

Please cite as:
ElSherief M, Sumner SA, Jones CM, Law RK, Kacha-Ochana A, Shieber L, Cordier L, Holton K, De Choudhury M
Characterizing and Identifying the Prevalence of Web-Based Misinformation Relating to Medication for Opioid Use Disorder: Machine Learning Approach
J Med Internet Res 2021;23(12):e30753
URL: https://www.jmir.org/2021/12/e30753
doi:10.2196/30753
PMID:34941555

Original Paper

Tracking Private WhatsApp Discourse About COVID-19 in Singapore: Longitudinal Infodemiology Study

Edina YQ Tan[1,2], BA; Russell RE Wee[3], BA; Young Ern Saw[1], BA; Kylie JQ Heng[1]; Joseph WE Chin[1], BA; Eddie MW Tong[3], PhD; Jean CJ Liu[1,2,4,5], PhD

[1]Division of Social Sciences, Yale-NUS College, Singapore, Singapore

[2]Centre for Sleep and Cognition, Yong Loo Lin School of Medicine, Singapore, Singapore

[3]Department of Psychology, National University of Singapore, Singapore, Singapore

[4]Neuroscience and Behavioral Disorders Program, Duke-NUS Medical School, Singapore, Singapore

[5]Centre for Trusted Internet and Community, National University of Singapore, Singapore, Singapore

Corresponding Author:
Jean CJ Liu, PhD
Division of Social Sciences
Yale-NUS College
28 College Ave West
Singapore, 138527
Singapore
Phone: 65 66013694
Email: jeanliu@yale-nus.edu.sg

Abstract

Background: Worldwide, social media traffic increased following the onset of the COVID-19 pandemic. Although the spread of COVID-19 content has been described for several social media platforms (eg, Twitter and Facebook), little is known about how such content is spread via private messaging platforms, such as WhatsApp (WhatsApp LLC).

Objective: In this study, we documented (1) how WhatsApp is used to transmit COVID-19 content, (2) the characteristics of WhatsApp users based on their usage patterns, and (3) how usage patterns link to COVID-19 concerns.

Methods: We used the experience sampling method to track day-to-day WhatsApp usage during the COVID-19 pandemic. For 1 week, participants reported each day the extent to which they had received, forwarded, or discussed COVID-19 content. The final data set comprised 924 data points, which were collected from 151 participants.

Results: During the weeklong monitoring process, most participants (143/151, 94.7%) reported at least 1 COVID-19–related use of WhatsApp. When a taxonomy was generated based on usage patterns, around 1 in 10 participants (21/151, 13.9%) were found to have received and shared a high volume of forwarded COVID-19 content, akin to super-spreaders identified on other social media platforms. Finally, those who engaged with more COVID-19 content in their personal chats were more likely to report having COVID-19–related thoughts throughout the day.

Conclusions: Our findings provide a rare window into discourse on private messaging platforms. Such data can be used to inform risk communication strategies during the pandemic.

Trial Registration: ClinicalTrials.gov NCT04367363; https://clinicaltrials.gov/ct2/show/NCT04367363

(J Med Internet Res 2021;23(12):e34218) doi:10.2196/34218

KEYWORDS

social media; WhatsApp; infodemiology; misinformation; COVID-19; tracking; surveillance; app; longitudinal; Singapore; characteristic; usage; pattern; well-being; communication; risk

Introduction

WhatsApp (WhatsApp LLC) is the most commonly used messaging app worldwide; it has 1.5 billion users across 180 countries [1]. On account of its large user base and near-instant message transmission capabilities, the platform has played a critical role in risk communication during the COVID-19 pandemic.

WhatsApp has been co-opted by government agencies and the World Health Organization to disseminate official COVID-19 updates [2]. However, while this showcases the platform's ability to reach a large sector of the population, this feature has also made it a vessel for misinformation. For example, at the beginning of the pandemic, WhatsApp noted a 40% surge in usage [3]. This was paired with a high volume of message forwarding activity that was widely believed to support misinformation. As a result, the platform restricted the number of individuals to whom a message could be forwarded simultaneously [4,5].

Despite these restrictions, a survey in India found that 1 in 2 participants had received COVID-19 misinformation through WhatsApp or Facebook [6]. Likewise, WhatsApp was identified by Hong Kong residents as the foremost source for COVID-19–related rumors [7]. As misinformation can jeopardize public health strategies, these findings underscore the need for infodemiological studies that document how COVID-19 content spreads through WhatsApp.

To date however, the bulk of infodemiology studies have focused on social media platforms in which content is publicly accessible (eg, Twitter and Facebook) [8,9]. In contrast, research on WhatsApp has proven to be elusive because of the platform's private nature; its end-to-end encryption software ensures that only senders and recipients have access to messages sent through the platform. Nonetheless, WhatsApp research remains a priority; aside from its popularity and role in disseminating crisis-related misinformation [10,11], insights from public posts are also unlikely to generalize to WhatsApp's private messages [12]. It thus remains unclear as to who sends COVID-19–related messages, who receives such messages, and what manner such messages are sent.

To address these gaps in the literature, we designed a study to (1) describe the base rate of COVID-19 content dissemination, (2) understand WhatsApp users, and (3) examine correlates of usage patterns. Specifically, we used the experience sampling method to track WhatsApp usage amid everyday routines across 1 week [13,14]. We asked participants to report each day their frequency of receiving, forwarding, or discussing COVID-19–related content. Through this method, we generated a taxonomy of participants based on their usage patterns and examined whether day-to-day variations in WhatsApp usage predicted COVID-19–related concerns.

Methods

Recruitment

From March 17 to May 7, 2020, participants were recruited from the general community via advertisements placed in Facebook and WhatsApp community groups (eg, residential groups, workplace groups, and university groups), posts on popular web-based forums, and paid Facebook advertisements targeting Singapore-based users. All study activities took place on the web-based survey platform Qualtrics (Qualtrics International Inc), and participants were reimbursed with SGD $5 (US $3.65) upon study completion. The study protocol was approved by the Yale-NUS (National University of Singapore) College Ethics Review Committee (protocol record: 2020-CERC-001) and was preregistered at ClinicalTrials.gov (trial number: NCT04367363).

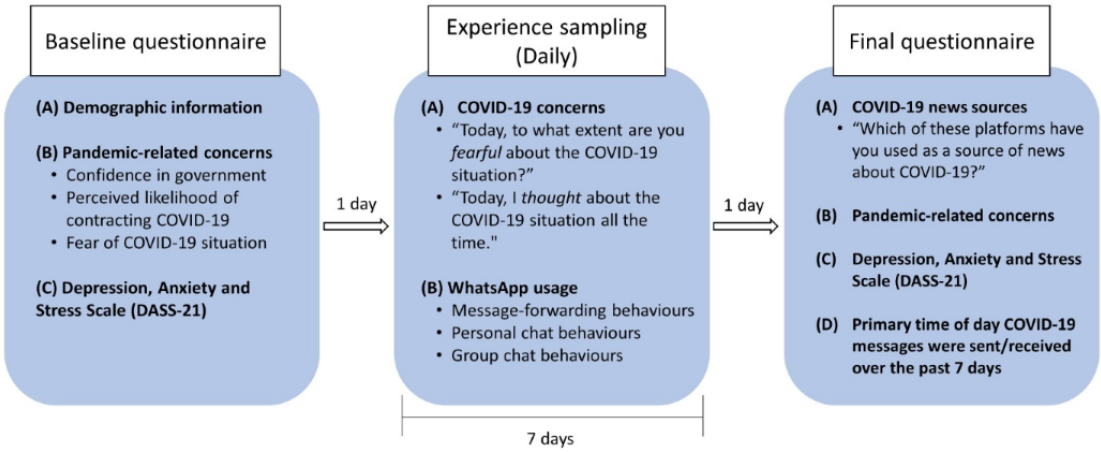
Participants

The participants were 151 adults who met the following inclusion criteria: (1) aged 21 years or older, (2) had lived in Singapore for at least 2 years, and (3) had a WhatsApp account.

Measures

Following the provision of informed consent, participants completed (1) a baseline questionnaire, (2) experience sampling responses daily for 7 days, and (3) a final questionnaire (Figure 1).

Figure 1. Schematic of study procedures. All participants completed a baseline questionnaire. This was followed by 7 days of experience sampling, during which participants addressed questions about COVID-19 concerns and WhatsApp usage daily. Participants completed a final questionnaire 1 day after the experience sampling procedure ended. DASS-21: 21-item Depression, Anxiety and Stress Scale.



Experience Sampling

As the primary form of data collection, we used the experience sampling method to capture COVID-19 chatter on WhatsApp. Through this method, we collected 924 data points across 151 participants (compliance rate: 924/1057, 87.4%).

For 7 days, participants accessed a web-based survey each evening (at 9:30 PM) to report their WhatsApp usage for the

day. Participants indicated whether they had forwarded messages related to COVID-19 (“yes” or “no”). We focused on message forwarding as a proxy indicator for high-risk content, since (1) large Twitter studies have observed that misinformation is more likely to be shared than posts that are true [15] and (2) WhatsApp developers had previously linked forwarded messages to misinformation [4,5]. If participants had forwarded COVID-19 content, they were then asked about (1) the number of unique COVID-19 messages they had forwarded and (2) the number of unique groups and individuals to which they had forwarded messages.

Participants were also asked about their personal chats (ie, their one-to-one chats on WhatsApp). They indicated whether COVID-19 messages had been forwarded to them in personal chats (“yes” or “no”). If so, they were asked about (1) the number of unique messages they had received and (2) the number of different people from which they had received messages. Thereafter, participants recounted whether they had discussed COVID-19 in conversations where either they or the other party generated messages related to COVID-19 (“yes” or “no”). If so, they were asked about how many unique chats were involved.

Finally, for group chats, participants were asked if COVID-19 had been mentioned in any of their WhatsApp groups by at least 1 other person (not including themselves; “yes” or “no”). This could have occurred either through others forwarding messages or through others generating their own comments. Affirmative responses were followed with a question on how many WhatsApp groups had done so.

Aside from WhatsApp metrics, participants also reported their COVID-19 concerns for the day; they were asked about (1) how afraid they felt about the COVID-19 situation (4-point scale: 1=“Not scared at all”; 4=“Very scared”) and (2) whether they thought about the COVID-19 situation all the time (5-point scale: 1=“Not at all true”; 5=“Very true”).

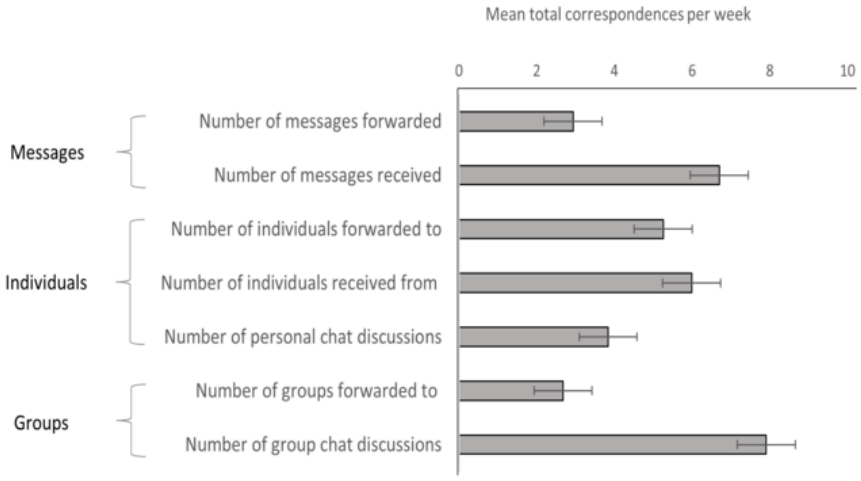
Baseline and Final Questionnaires

To characterize the participants, we included baseline and final questionnaires in which participants reported demographics (age, gender, religion, ethnicity, marital status, education, house type, household size, citizenship, country of birth, and number of years in Singapore), the time of day when they read and sent COVID-19 messages on WhatsApp (mostly in the morning, afternoon, evening, or late night or throughout the day), and sources through which they obtained COVID-19 news (eg, printed newspapers, radio, WhatsApp, and YouTube). Additionally, participants completed the 21-item Depression, Anxiety and Stress Scale (DASS-21) [16] to evaluate their mental health during the pandemic. Participants were also asked about their responses to the pandemic [2,17], that is, (1) how confident they were that the government could control the nationwide spread of COVID-19 (1=“Not confident at all; 4=“Very confident”), (2) their perceptions on how likely that they or someone in their immediate household would contract COVID-19 (1=“Not at all likely”; 4=“Very likely”), and (3) how fearful they were about the situation in the country (1=“Not scared at all”; 4=“Very scared”).

Statistical Analysis

First, we summarized the data as counts with percentages or as means with SDs, focusing on the following seven quantitative WhatsApp usage variables (Figure 2): the number of (1) COVID-19 messages that participants forwarded, (2) groups to whom messages were forwarded, (3) individuals to whom messages were forwarded, (4) forwarded messages received, (5) individuals from whom messages were received, (6) personal chats involving COVID-19–related conversations, and (7) group chats discussing COVID-19. Multimedia Appendix 1 shows the pattern of correlations across these variables.

Figure 2. Distribution of COVID-19–related behaviors on WhatsApp. In a weeklong experience sampling procedure, participants reported the extent to which they engaged in COVID-19–related behaviors on WhatsApp (either by forwarding or receiving messages or in conversations). Horizontal bars represent the total amount of each activity captured (averaged across all participants). Horizontal lines represent the 95% CIs for the means.



Second, to understand WhatsApp user profiles, we performed a latent profile analysis to create a taxonomy of participants based on their WhatsApp usage (R package *mclust* [18]). Latent profile analysis is a bottom-up statistical clustering method for defining classes of people based on common characteristics. By using all observations of a continuous dependent variable, classes are created such that within each class, indicator variables are statistically uncorrelated [19]. We thus used this

technique to cluster participants based on their responses to the seven WhatsApp usage variables; values were obtained by aggregating the reported frequency of each variable over the week. To uncover clusters, we used Gaussian mixture models and assigned participants to clusters by using Bayesian probabilities. The final number of clusters was determined by using the Bayesian information criterion, the integrated completed likelihood criterion, and a bootstrap likelihood test.

Finally, we examined whether day-to-day variations in COVID-19 WhatsApp chatter could be tracked based on variations in COVID-19 concerns. We quantified COVID-19 chatter on personal and group chats, so that such chatter could be used as predictors. For personal chats, the following variables were summed for each day and for each participant: the number of (1) individuals to whom COVID-19 messages were forwarded, (2) individuals from whom forwarded messages were received, and (3) personal conversations discussing COVID-19. For group chats, the following variables were summed: the number of (1) groups participants to whom COVID-19 messages were forwarded and (2) groups where COVID-19 messages were mentioned. Scores were grand-mean centered by subtracting the mean number of chats across subjects and time points from each score (number of chats: mean 2.47 and mean 1.29, respectively). In addition, we created between- and within-subject versions of each predictor [20]. The final analyses involved linear mixed-effects models for each outcome measure (fear of and thoughts about COVID-19). The following were entered as fixed effects: time (centered such that 0 referred to the middle of the week), daily personal chats (between subjects), daily personal chats (within subjects), daily group chats (between subjects), and daily group chats (within subjects). Random intercepts accounted for correlated data resulting from repeated measures.

Across all analyses, the type 1 decision-wise error rate was controlled at an α of .05. All statistical analyses were conducted in R 3.5.0 (R Foundation for Statistical Computing) and SPSS 25 (IBM Corporation).

Results

Baseline Participant Characteristics

As shown in Table 1, 68.9% (104/151) of participants were female, and their mean age was 36.35 (SD 14.7) years. Participants were predominantly of Asian ethnicity (Chinese: 140/151, 92.7%) and had at least a postsecondary education (133/151, 88.1%). Further, 39.7% (60/151) of participants were married, and the majority (105/151, 69.5%) belonged to households of at least 4 members.

Table 1. Participant characteristics as a function of COVID-19 WhatsApp usage patterns.

Characteristic	Chronic users (n=21)	Receiving users (n=47)	Discursive users (n=46)	Minimal users (n=37)	All participants (N=151)
Age (years), mean (SD)	44.1 (14.5)	41.0 (15.5)	29.7 (10.7)	34.4 (14.5)	36.35 (14.70)
Gender, n (%)					
Female	13 (62)	34 (72)	29 (63)	28 (76)	104 (69)
Male	8 (38)	13 (28)	17 (37)	9 (24)	47 (31)
Ethnicity , n (%)					
Chinese	20 (95)	42 (89)	42 (91)	36 (97)	140 (93)
Indian	0 (0)	2 (5)	3 (7)	0 (0)	5 (3)
Malay	0 (0)	2 (5)	1 (2)	0 (0)	3 (2)
Other	1 (5)	1 (1)	0 (0)	1 (3)	3 (2)
Religion, n (%)					
Christianity (Protestant)	8 (38)	17 (36)	16 (35)	13 (35)	54 (36)
No religion	3 (14)	14 (30)	11 (24)	10 (27)	38 (25)
Buddhism	4 (19)	9 (19)	8 (18)	11 (30)	32 (21)
Roman Catholicism	4 (19)	4 (9)	6 (13)	2 (5)	16 (11)
Taoism or Chinese traditional beliefs	1 (5)	0 (0)	2 (4)	1 (3)	4 (3)
Islam	1 (5)	3 (6)	1 (2)	0 (0)	5 (3)
Hinduism	0 (0)	0 (0)	2 (4)	0 (0)	2 (1)
Marital status , n (%)					
Married	13 (62)	24 (51)	8 (17)	15 (41)	60 (40)
Single	6 (28)	15 (32)	25 (55)	12 (32)	58 (38)
Dating	1 (5)	7 (15)	12 (26)	9 (24)	29 (19)
Widowed, separated, or divorced	1 (5)	1 (2)	0 (0)	1 (3)	3 (2)
Did not answer	0 (0)	0 (0)	1 (2)	0 (0)	1 (1)
Educational level, n (%)					
O level	1 (5)	4 (9)	1 (2)	6 (16)	12 (8)
Junior college	2 (10)	5 (10)	9 (19)	9 (24)	25 (17)
Institute of Technical Education	1 (5)	1 (2)	1 (2)	0 (0)	3 (2)
Polytechnic or diploma	2 (10)	13 (28)	7 (15)	4 (11)	26 (17)
University (undergraduate)	11 (51)	21 (45)	21 (46)	16 (43)	69 (46)
University (postgraduate)	4 (19)	1 (2)	3 (7)	2 (6)	10 (7)
Other	0 (0)	2 (4)	3 (7)	0 (0)	5 (3)
Did not answer	0 (0)	0 (0)	1 (2)	0 (0)	1 (1)
House type, n (%)					
HDB[a] flat (1-2 rooms)	0 (0)	0 (0)	0 (0)	1 (3)	1 (1)
HDB flat (3 rooms)	0 (0)	2 (4)	2 (4)	2 (5)	6 (4)
HDB flat (4 rooms)	2 (10)	9 (19)	10 (22)	10 (27)	31 (21)
HDB flat (5 rooms)	3 (14)	19 (40)	14 (31)	11 (30)	47 (31)
Condominium	12 (57)	12 (26)	11 (24)	10 (27)	45 (30)
Landed property	4 (19)	4 (9)	7 (15)	2 (5)	17 (11)
Did not answer	0 (0)	1 (2)	2 (4)	1 (3)	4 (3)
Household size (number of members), n (%)					

Characteristic	Chronic users (n=21)	Receiving users (n=47)	Discursive users (n=46)	Minimal users (n=37)	All participants (N=151)
1	2 (10)	1 (2)	3 (7)	0 (0)	6 (4)
2	0 (0)	5 (11)	3 (7)	3 (8)	11 (7)
3	7 (33)	8 (17)	5 (11)	8 (22)	28 (19)
4	7 (33)	18 (38)	21 (45)	15 (40)	61 (40)
≥5	5 (24)	15 (32)	13 (28)	11 (30)	44 (29)
Did not answer	0 (0)	0 (0)	1 (2)	0 (0)	1 (1)
Citizenship, n (%)					
Singapore	18 (86)	46 (98)	42 (91)	36 (97)	142 (94)
Other	3 (14)	1 (2)	4 (9)	1 (3)	9 (6)
Country of birth, n (%)					
Singapore	17 (81)	45 (96)	38 (83)	33 (89)	133 (88)
Other	4 (19)	2 (4)	8 (17)	4 (11)	18 (12)
Number of years in Singapore, mean (SD)	39.67 (15.22)	39.60 (16.69)	26.43 (10.83)	31.65 (14.47)	33.65 (15.32)
21-item Depression, Anxiety and Stress Scale scores, mean (SD)					
Stress	9.52 (7.12)	8.61 (7.08)	9.56 (10.13)	10.81 (8.72)	9.57 (8.47)
Anxiety	4.38 (5.28)	5.13 (5.44)	5.33 (6.59)	5.89 (7.71)	5.28 (6.36)
Depression	8.10 (6.52)	7.22 (6.86)	9.47 (9.73)	10.76 (9.54)	8.90 (8.50)
Pandemic-related concerns (score), mean (SD)					
Fear of COVID-19 situation	2.29 (0.46)	2.53 (0.65)	2.22 (0.74)	2.27 (0.69)	2.34 (0.67)
Confidence in government	3.33 (0.58)	3.23 (0.63)	3.29 (0.66)	3.24 (0.72)	3.27 (0.65)
Perceived likelihood of contracting COVID-19	2.71 (0.64)	2.74 (0.53)	2.78 (0.56)	2.76 (0.60)	2.75 (0.57)

[a]HDB: housing and development.

Base Rate of COVID-19 WhatsApp Usage

Participants' self-reports revealed that WhatsApp was the second most common source for COVID-19 news after news websites or apps (Figure 3). By quantifying this through 1 week of experience sampling, we found that nearly all participants (143/151, 94.7%; 95% CI 90-98%) reported at least 1 COVID-19 related use of WhatsApp. Namely, around 1 in 2 participants (79/151, 52.3%; 95% CI 44%-60%) forwarded at least 1 COVID-19 message (to either individuals or groups), 78.1% (118/151; 95% CI 71%-84%) received at least 1 forwarded message in personal chats, 66.2% (100/151; 95% CI 58%-74%) engaged in personal chat conversations about COVID-19, and 88.1% (133/151; 95% CI 82%-93%) had been in groups where COVID-19 was mentioned.

Figure 3. Sources of COVID-19 news. In a questionnaire, participants self-reported the sources from which they received COVID-19 news.

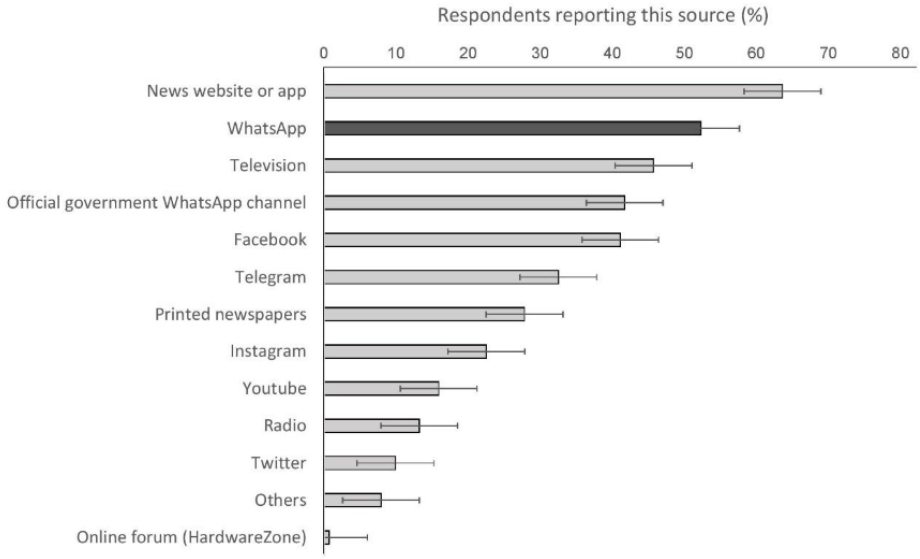


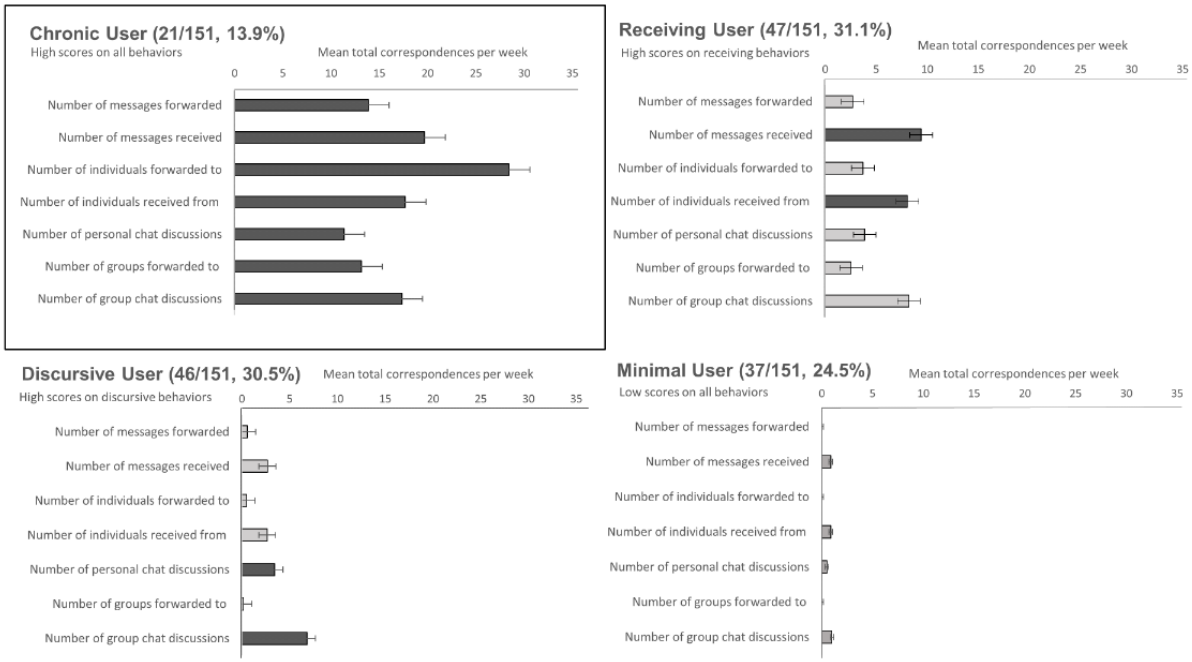
Figure 2 shows the extent to which participants engaged in each of these activities. On average, participants (1) received 2.3 times more messages than they forwarded and (2) were more likely to forward messages to individuals than to groups (average of 5.3 messages per week vs 2.7 messages per week, respectively). Beyond passive engagement, participants also took part in an average of 3.8 one-to-one conversations about COVID-19 during the week; however, these interactions occurred less frequently than the sending or receiving of forwarded messages in group chats.

Characterizing Participants Based on COVID-19 WhatsApp Usage

Latent Profile Analysis: Generating a Taxonomy of WhatsApp Usage

Although most participants (143/151, 94.7%) received and shared COVID-19 content on WhatsApp, there were individual differences in usage patterns (Figure 4). Correspondingly, we conducted a latent profile analysis to understand how usage patterns clustered.

Figure 4. Taxonomy of COVID-19–related WhatsApp usage. By using latent profile analysis, we classified participants based on how they had used WhatsApp to engage with COVID-19 content during 1 week of monitoring. The figure depicts the WhatsApp usage activities of chronic users (top left), receiving users (top right), discursive users (bottom left), and minimal users (bottom right). Horizontal lines represent the 95% CIs for the means.



A 4-cluster solution yielded the lowest absolute Bayesian information criterion values (Multimedia Appendix 2), resulting in the following taxonomy (Figure 4). First, 13.9% (21/151) of participants were chronic users, who exhibited high levels of activity with regard to each of the WhatsApp usage variables. Correspondingly, this group of participants was responsible for receiving and transmitting a large volume of forwarded COVID-19 messages; they sent the messages both to individual

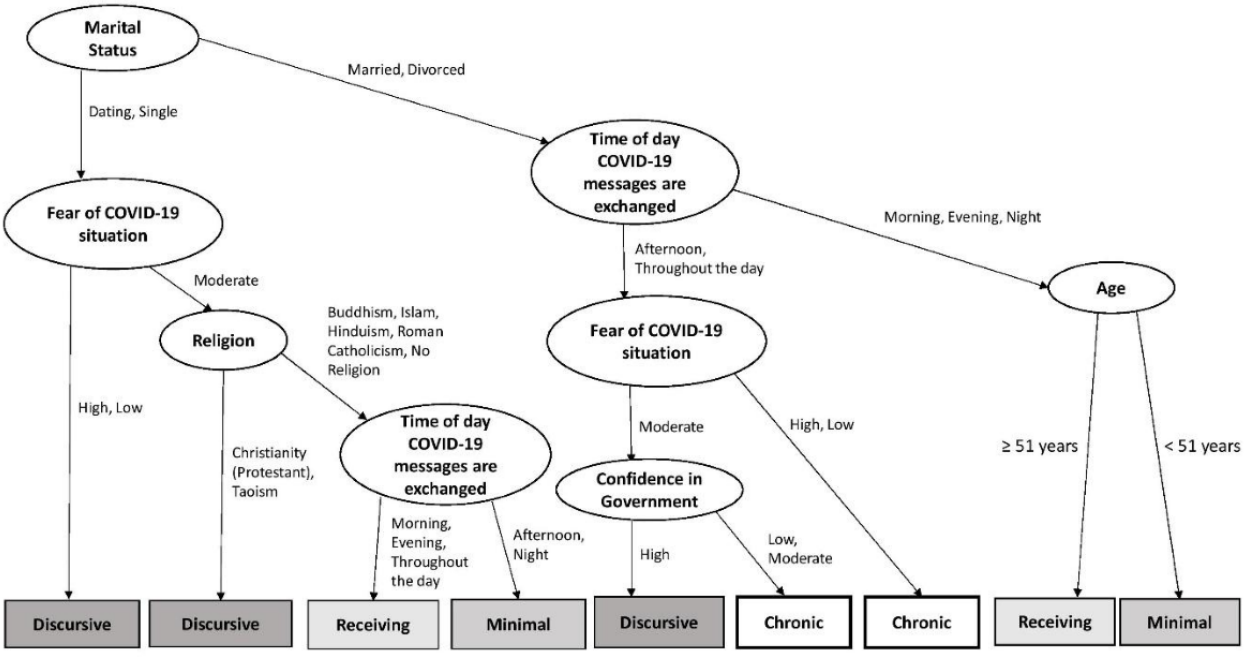
contacts and to groups. Second, 31.1% (47/151) of participants were receiving users, who were distinguished by their receipt of multiple forwarded COVID-19 messages. Although this group discussed COVID-19 frequently in group chats, they rarely passed along the forwarded COVID-19 messages that they had received. A third group – discursive users (46/151, 30.5%) – had low exposure to forwarded COVID-19 messages and primarily engaged with COVID-19 content through personal and group chats. Finally, 24.5% (37/151) of participants were minimal users, who had low levels of engagement with COVID-19 content overall.

Understanding User Characteristics

As an exploratory analysis, we performed a classification tree analysis to predict WhatsApp user types based on demographics, COVID-19 concerns, depression and anxiety scores (DASS-21), and the time of day when participants used WhatsApp. We performed recursive partitioning (*rpart*)—a machine learning technique that allows multiple variables to be analyzed simultaneously and supports the modeling of complex, nonlinear relations among predictors [21]. To avoid overfitting, the final tree was pruned by selecting the tree size with the lowest cross-validation error, which, for our data set, was a tree size of 8.

As shown in Figure 5, chronic users were more likely to be married or divorced and more likely to send messages either throughout the day or in the afternoon. In terms of responses to the COVID-19 pandemic, chronic users either (1) had extreme fears of the COVID-19 situation (low or high levels of fear) or (2) had moderate fears paired with low confidence in the government's response (low or moderate confidence in government).

Figure 5. Classification tree analysis. Recursive partitioning was used to predict which of the four WhatsApp usage profiles (chronic, receiving, discursive, or minimal) participants belonged to based on baseline questionnaire measures (demographics; COVID-19 concerns; scores on the 21-item Depression, Anxiety and Stress Scale; and time of WhatsApp usage). The final tree model is presented as a flowchart; factors are chosen at each level to categorize the maximal number of participants. Marital status, the time of WhatsApp usage, and age emerged as the primary predictors (model classification accuracy: 64.2%; above the chance level of 25%).



Discursive users were more likely to be single or be dating and had either (1) extreme levels of COVID-19–related fears (either high or low levels of fear) or (2) moderate fear levels alongside Christian or Taoist affiliations. A subgroup of discursive users were, like chronic users, married or divorced and had moderate levels of COVID-19–related fears. However, they were distinguished from chronic users based on their high confidence in the government (chronic users had lower confidence in the government).

Finally, receiving, and minimal users had similar profiles. If they were single or were dating, both sets of users tended to have moderate levels of COVID-19–related fears, had a wide range of religious backgrounds, and were distinguished based on the time of day when they received COVID-19–related messages (receiving users: in the morning, in the evening, and throughout the day; minimal users: in the afternoon and at night). If they were married or divorced, both sets of users tended to send messages at only 1 time of the day (morning, evening, or night) and were distinguished based on age (receiving users: aged ≥51 years; minimal users: aged <51 years). Table 1 describes the demographic characteristics of the four user profiles.

Does WhatsApp Usage Relate to COVID-19 Concerns?

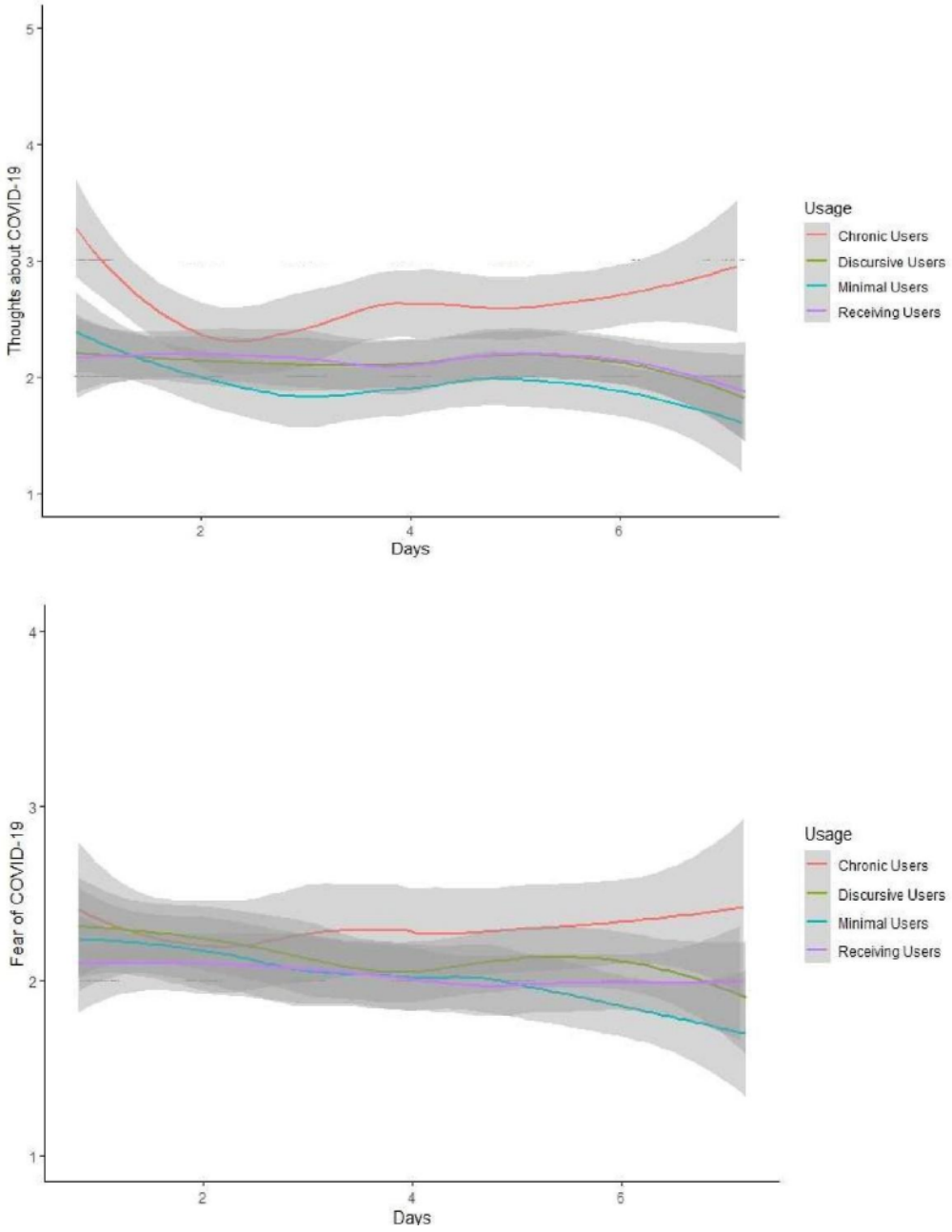
Finally, we conducted linear mixed-effects models to examine whether WhatsApp usage was related to COVID-19 concerns (Table 2). As shown in Figure 6, day-to-day COVID-19–related fears and thoughts fluctuated (fears: $t_{249.13}=-3.72$; $P<.001$; thoughts: $t_{297.02}=-2.36$; $P=.02$).

Table 2. Parameter estimates for the multi-level model of thoughts about COVID-19 (model 1) and the fear of COVID-19 (model 2) as a function of participants' daily WhatsApp use (personal chats and group chats).

Model and effects	Estimate, β (SE; 95% CI)	*t* test[a]*(df)* or *Z*	*P* value
Model 1 outcome: thoughts about COVID-19			
Fixed effects			
Intercept	2.18 (0.07; 2.05 to 2.31)	32.81 (135.68)	<.001
Time (centered)	−.03 (.01; −.05 to 0)	−2.36 (297.02)	.02
Daily personal chat usage (between subjects)	.04 (.02; 0 to .07)	2.36 (164.48)	.02
Daily personal chat usage (within subjects)	0 (.01; −.01 to .02)	0.42 (17.63)	.68
Daily group chat usage (between subjects)	.05 (.06; −.06 to .17)	0.89 (141.17)	.37
Daily group chat usage (within subjects)	0 (.03; −.06 to .05)	−0.08 (14.09)	.93
Random effects			
Intercept (between subjects)	.56 (.08; .42 to .75)	6.89	<.001
Residual (within subjects)	.37 (.02; .33 to .43)	14.90	<.001
Autocorrelation (within subjects)	.24 (.05; .14 to .33)	4.97	<.001
Model 2 outcome: fear of COVID-19			
Fixed effects			
Intercept	2.10 (0.06; 1.98 to 2.21)	36.37 (144.90)	<.001
Time (centered)	−.03 (.01; −.05 to −.02)	−3.72 (249.13)	<.001
Daily personal chat usage (between subjects)	.01 (.01; −.02 to .04)	0.85 (155.44)	.39
Daily personal chat usage (within subjects)	.01 (.01; −.01 to .02)	1.22 (24.97)	.24
Daily group chat usage (between subjects)	.02 (.05; −.07 to .12)	0.49 (128.59)	.62
Daily group chat usage (within subjects)	−.03 (.02; −.06 to .01)	−1.42 (28.88)	.17
Random effects			
Intercept (between subjects)	.44 (.06; .34 to .58)	7.47	<.001
Residual (within subjects)	.21 (.01; .18 to .23)	13.83	<.001
Autocorrelation (within subjects)	.26 (.05; to .16 to .35)	5.16	<.001

[a]The *t* test was 2-tailed.

Figure 6. COVID-19–related thoughts and fears over 1 week. Day-to-day variations in COVID-19–related thought (top) and fear levels (bottom) as a function of WhatsApp user profiles. The shaded grey areas represent 95% CIs.



For thoughts about COVID-19, there was a significant effect in WhatsApp personal chat usage at a between-subjects level ($t_{164.48}$=2.36; *P*=.02), that is, participants who handled larger amounts of COVID-19 content in their personal chats reported having more thoughts about COVID-19 (relative to participants who handled smaller amounts of COVID-19 content). However, the corresponding effect for group chats was not significant ($t_{141.17}$=0.89; *P*=.37). At the within-subjects level, neither day-to-day fluctuations in personal chat activities nor those in group chat activities significantly predicted thoughts about COVID-19 (smallest *P*=.68).

For COVID-19–related fears, we found no significant effects in any of the WhatsApp usage variables (smallest *P*=.17). For sensitivity analyses, we repeated both models and used group membership as a fixed factor in place of personal and group chat usage, and our primary conclusions did not change (Multimedia Appendix 3).

Discussion

Principal Findings

The ongoing pandemic has drawn attention to the role of social media in public health. Against this backdrop, we present the first infodemiological study to document the spread of COVID-19 content through WhatsApp. By tracking daily WhatsApp usage for 1 week, we found that (1) nearly every participant engaged in COVID-19 chatter and (2) participants were more likely to share or receive forwarded messages than to engage in conversations about COVID-19.

The volume of forwarded messages that we observed raises concern. On other social media platforms, forwarding behaviors have been linked to the spread of misinformation. For example, a study of 4.5 million Twitter posts found that misinformation was 70% more likely to be shared than posts that were true; correspondingly, any single retweet had a higher probability of containing false news rather than truthful news [15]. Although analogous research has not been conducted on WhatsApp, the app's developers have likewise deemed forwarded messages as a high-risk source of misinformation [4,5].

With regard to the extent that forwarded messages carry misinformation [1,2], our latent profile analyses revealed that around 1 in 10 (21/151, 13.9%) participants were chronic users who received and shared a large volume of such messages. Notably, chronic users disseminated an average of 14 forwarded messages during the week, which is approximately 5 times the number of messages that were sent by all participants in this study. This is reminiscent of research on other social media platforms (eg, Twitter) where a small group of super sharers and super consumers are responsible for the bulk of shared misinformation [22]. Given the potential influence of this group, further research is needed to understand (1) the profile of chronic users, (2) the reasons why they forward messages, and (3) how their forwarding activities may influence outcomes during health crises.

Aside from chronic users, our study also found that around 1 in 3 (47/151, 31.1%) participants were receiving users who had high exposure to forwarded COVID-19 content. Receiving users tended to be older (in line with misinformation studies on Facebook [23]) but were otherwise moderate in terms of their profiles, that is, in terms of COVID-19–related fears or religion (they came from a diverse religious background). Although this group did not spread forwarded messages themselves, their high exposure may nonetheless leave them susceptible to false beliefs. Correspondingly, we urge researchers to conduct further research to understand how WhatsApp use among receiving users influences health behaviors.

Finally, we also found that WhatsApp users who discussed COVID-19 in their personal chats were more likely to think about COVID-19 throughout the day. As similar forms of rumination (involving frequent and persistent thoughts) have been linked to clinical depression [24], this finding may implicate COVID-19 chatter as a risk factor for poorer well-being [25]. Future studies should thus explore this possibility and the potential mechanisms involved.

Implications

Taken together, our findings on WhatsApp message transmission have several implications for public health responses during a crisis. First, our taxonomy of user profiles provides a basis for targeted risk communication. Our findings suggest that public health agencies may need to reach out proactively to chronic and receiving users, who handle the bulk of forwarded COVID-19 content on WhatsApp. One possible intervention may be encouraging these users to subscribe to official WhatsApp channels for updates (eg, updates from the World Health Organization) [2] to capitalize on their pre-existing readiness to use the platform.

Tracking WhatsApp chatter may also result in new opportunities for detecting disease outbreaks. The nascent field of digital epidemiology seeks to model how diseases spread by monitoring digital data sources (eg, through Google search data and Twitter posts) [26]. Although the content of private WhatsApp messages is difficult to track, the volume or nature of messages (eg, forwarded messages) may provide information that can be used to support disease surveillance. To this end, further research is needed to explore the predictive utility of message transmission dynamics on WhatsApp.

Limitations

In reporting these findings, we noted several limitations. First, we opted to study WhatsApp—the most widely used messaging app. At this juncture, it is unclear whether our results generalize to other messaging apps (eg, Facebook Messenger and Telegram).

Second, our recruitment strategy was limited by the nature of the pandemic. Owing to infectious disease protocols and the short time period when the amount of crisis-related communication was high [27], our data collection process was restricted in terms of the recruitment strategy (web-based sampling), sample size (151 participants), and time frame (1 week per participant). Further research is needed to examine whether our findings generalize to the broader population.

Finally, although the experience sampling method captured WhatsApp usage in participants' naturalistic settings, the method nonetheless required self-reports. By extending our findings, future studies will benefit from having objective metrics of WhatsApp usage.

Conclusions

In conclusion, we used the experience sampling method to capture COVID-19 chatter on WhatsApp for the first time. In total, we tracked 924 days' worth of chatter in situ, revealing (1) the sheer prevalence of WhatsApp usage, (2) a typology of WhatsApp users, and (3) a link between usage patterns and constant thoughts about the pandemic. These findings have implications for health communication and disease surveillance, bringing the field 1 step closer to characterizing WhatsApp usage and using these data to gain insights on individual and societal concerns.

Acknowledgments

This research was funded by a grant awarded to JCJL from the NUS (National University of Singapore) Centre for Trusted Internet and Community (grant CTIC-RP-20-09). The first author's (EYQT) involvement was funded by a center grant awarded to the Centre for Sleep and Cognition.

Authors' Contributions

RREW, KJQH, and JWEC contributed to the study design and collected the data. EYQT and YES analyzed the data and wrote the first draft of the manuscript. JCJL and EMWT conceived the study, had oversight of the overall project, and supervised both the data analyses and manuscript preparation. All authors provided feedback on the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation matrix of the seven quantitative WhatsApp usage variables.
[DOCX File , 179 KB - jmir_v23i12e34218_app1.docx]

Multimedia Appendix 2
Model fit indices for latent profile analysis. Lower absolute values indicate better model fit.
[DOCX File , 197 KB - jmir_v23i12e34218_app2.docx]

Multimedia Appendix 3
Model parameter estimates (group membership). Parameter estimates for the multi-level model of thoughts about COVID-19 (model 1) and the fear of COVID-19 (model 2) as a function of participants' group memberships.
[DOCX File , 30 KB - jmir_v23i12e34218_app3.docx]

References

1. Singh M. WhatsApp hits 2 billion users, up from 1.5 billion 2 years ago. TechCrunch. 2020 Feb 12. URL: http://techcrunch.com/2020/02/12/whatsapp-hits-2-billion-users-up-from-1-5-billion-2-years-ago/ [accessed 2020-12-01]
2. Liu JCJ, Tong EMW. The relation between official WhatsApp-distributed COVID-19 news exposure and psychological symptoms: Cross-sectional survey study. J Med Internet Res 2020 Sep 25;22(9):e22142 [FREE Full text] [doi: 10.2196/22142] [Medline: 32877349]
3. COVID-19 barometer: Consumer attitudes, media habits and expectations. Kantar. 2020 Apr 03. URL: https://www.kantar.com/Inspiration/Coronavirus/COVID-19-Barometer-Consumer-attitudes-media-habits-and-expectations [accessed 2020-12-01]
4. Newton C. WhatsApp puts new limits on the forwarding of viral messages. The Verge. 2020 Apr 07. URL: https://www.theverge.com/2020/4/7/21211371/whatsapp-message-forwarding-limits-misinformation-coronavirus-india [accessed 2020-12-01]
5. Singh M. WhatsApp's new limit cuts virality of 'highly forwarded' messages by 70%. TechCrunch. 2020 Apr 27. URL: http://techcrunch.com/2020/04/27/whatsapps-new-limit-cuts-virality-of-highly-forwarded-messages-by-70/ [accessed 2020-12-01]
6. Bapaye JA, Bapaye HA. Demographic factors influencing the impact of coronavirus-related misinformation on WhatsApp: Cross-sectional questionnaire study. JMIR Public Health Surveill 2021 Jan 30;7(1):e19858 [FREE Full text] [doi: 10.2196/19858] [Medline: 33444152]
7. Luk TT, Zhao S, Weng X, Wong JYH, Wu YS, Ho SY, et al. Exposure to health misinformation about COVID-19 and increased tobacco and alcohol use: a population-based survey in Hong Kong. Tob Control 2021 Nov;30(6):696-699. [doi: 10.1136/tobaccocontrol-2020-055960] [Medline: 32855353]
8. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
9. 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]
10. Gireesh KV, Dineshan K. Social media misinformation in disaster situations: A case of 2018 Kerala flood. Library Progress (International) 2019;39(2):289-296. [doi: 10.5958/2320-317x.2019.00032.1]
11. Simon T, Goldberg A, Leykin D, Adini B. Kidnapping WhatsApp – Rumors during the search and rescue operation of three kidnapped youth. Comput Human Behav 2016 Nov;64:183-190 [FREE Full text] [doi: 10.1016/j.chb.2016.06.058]
12. Lottridge D, Bentley FR. Let's hate together: How people share news in messaging, social, and public networks. 2018 Apr Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Quebec, Canada p. 1-13. [doi: 10.1145/3173574.3173634]
13. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: Csikszentmihalyi M, editor. Flow and the Foundations of Positive Psychology. Dordrecht, Netherlands: Springer; 2014:35-54.

14. Hektner JM, Schmidt JA, Csikszentmihalyi M. Experience Sampling Method: Measuring the Quality of Everyday Life. Thousand Oaks, California: SAGE Publications Inc; 2007.
15. 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]
16. Lovibond PF, Lovibond SH. 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] [Medline: 7726811]
17. Mesch GS, Schwirian KP. Vaccination hesitancy: fear, trust, and exposure expectancy of an Ebola outbreak. Heliyon 2019 Jul 18;5(7):e02016 [FREE Full text] [doi: 10.1016/j.heliyon.2019.e02016] [Medline: 31367685]
18. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. R J 2016 Aug;8(1):289-317 [FREE Full text] [Medline: 27818791]
19. Wade TD, Crosby RD, Martin NG. Use of latent profile analysis to identify eating disorder phenotypes in an adult Australian twin cohort. Arch Gen Psychiatry 2006 Dec;63(12):1377-1384. [doi: 10.1001/archpsyc.63.12.1377] [Medline: 17146012]
20. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY: Guilford Press; 2013.
21. Therneau TM, Atkinson EJ, Mayo Foundation. An introduction to recursive partitioning using the RPART routines. Mayo Clinic. 1997 Sep 03. URL: https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699 [accessed 2021-06-21]
22. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 U.S. presidential election. Science 2019 Jan 25;363(6425):374-378. [doi: 10.1126/science.aau2706] [Medline: 30679368]
23. Guess A, Nagler J, Tucker J. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Sci Adv 2019 Jan 09;5(1):eaau4586 [FREE Full text] [doi: 10.1126/sciadv.aau4586] [Medline: 30662946]
24. Moulds ML, Kandris E, Starr S, Wong ACM. The relationship between rumination, avoidance and depression in a non-clinical sample. Behav Res Ther 2007 Feb;45(2):251-261. [doi: 10.1016/j.brat.2006.03.003] [Medline: 16631110]
25. Blabst N, Diefenbach S. WhatsApp and wellbeing: A study on WhatsApp usage, communication quality and stress. https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2017.85; 2017 Jul Presented at: The 31st International BCS Human Computer Interaction Conference (HCI 2017) (HCI); July 3-6, 2017; Sunderland, UK. [doi: 10.14236/ewic/hci2017.85]
26. Tarkoma S, Alghnam S, Howell MD. Fighting pandemics with digital epidemiology. EClinicalMedicine 2020 Sep;26:100512 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100512] [Medline: 32864592]
27. Nsoesie EO, Cesare N, Müller M, Ozonoff A. COVID-19 misinformation spread in eight countries: Exponential growth modeling study. J Med Internet Res 2020 Dec 15;22(12):e24425 [FREE Full text] [doi: 10.2196/24425] [Medline: 33264102]

Abbreviations

DASS-21: 21-item Depression, Anxiety and Stress Scale
NUS: National University of Singapore

Edited by L Westafer, G Eysenbach; submitted 13.10.21; peer-reviewed by P Schulz, O Leal Neto; comments to author 24.11.21; revised version received 25.11.21; accepted 27.11.21; published 23.12.21.

Please cite as:
Tan EYQ, Wee RRE, Saw YE, Heng KJQ, Chin JWE, Tong EMW, Liu JCJ
Tracking Private WhatsApp Discourse About COVID-19 in Singapore: Longitudinal Infodemiology Study
J Med Internet Res 2021;23(12):e34218
URL: https://www.jmir.org/2021/12/e34218
doi:10.2196/34218
PMID:34881720

Original Paper

Investigating Individuals' Perceptions Regarding the Context Around the Low Back Pain Experience: Topic Modeling Analysis of Twitter Data

Robert[1*]; Pari Delir Haghighi[1*], PhD; Frada Burstein[1*], PhD; Donna Urquhart[2*], PhD; Flavia Cicuttini[2*], PhD, MD

[1]Department of Human-Centred Computing, Faculty of Information Technology, Monash University, Caulfield East, Australia

[2]Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[*]all authors contributed equally

Corresponding Author:
Pari Delir Haghighi, PhD
Department of Human-Centred Computing
Faculty of Information Technology
Monash University
Building H, Level 6
900 Dandenong Road
Caulfield East, 3145
Australia
Phone: 61 99032355
Email: pari.delirhaghighi@monash.edu

Abstract

Background: Low back pain (LBP) remains the leading cause of disability worldwide. A better understanding of the beliefs regarding LBP and impact of LBP on the individual is important in order to improve outcomes. Although personal experiences of LBP have traditionally been explored through qualitative studies, social media allows access to data from a large, heterogonous, and geographically distributed population, which is not possible using traditional qualitative or quantitative methods. As data on social media sites are collected in an unsolicited manner, individuals are more likely to express their views and emotions freely and in an unconstrained manner as compared to traditional data collection methods. Thus, content analysis of social media provides a novel approach to understanding how problems such as LBP are perceived by those who experience it and its impact.

Objective: The objective of this study was to identify contextual variables of the LBP experience from a first-person perspective to provide insights into individuals' beliefs and perceptions.

Methods: We analyzed 896,867 cleaned tweets about LBP between January 1, 2014, and December 31, 2018. We tested and compared latent Dirichlet allocation (LDA), Dirichlet multinomial mixture (DMM), GPU-DMM, biterm topic model, and nonnegative matrix factorization for identifying topics associated with tweets. A coherence score was determined to identify the best model. Two domain experts independently performed qualitative content analysis of the topics with the strongest coherence score and grouped them into contextual categories. The experts met and reconciled any differences and developed the final labels.

Results: LDA outperformed all other algorithms, resulting in the highest coherence score. The best model was LDA with 60 topics, with a coherence score of 0.562. The 60 topics were grouped into 19 contextual categories. "Emotion and beliefs" had the largest proportion of total tweets (157,563/896,867, 17.6%), followed by "physical activity" (124,251/896,867, 13.85%) and "daily life" (80,730/896,867, 9%), while "food and drink," "weather," and "not being understood" had the smallest proportions (11,551/896,867, 1.29%; 10,109/896,867, 1.13%; and 9180/896,867, 1.02%, respectively). Of the 11 topics within "emotion and beliefs," 113,562/157,563 (72%) had negative sentiment.

Conclusions: The content analysis of tweets in the area of LBP identified common themes that are consistent with findings from conventional qualitative studies but provide a more granular view of individuals' perspectives related to LBP. This understanding has the potential to assist with developing more effective and personalized models of care to improve outcomes in those with LBP.

(J Med Internet Res 2021;23(12):e26093) doi:10.2196/26093

KEYWORDS

low back pain; Twitter; content analysis; social media; topic modeling; patient-centered approach; pain experience; context of pain

Introduction

Low back pain (LBP) is the leading cause of disability worldwide [1,2]. Approximately 50%-80% of adults experience LBP at least once in their lives [3] and it is a leading cause of work absence and limits physical activities, posing a large economic burden [1,4]. In the United States, the total cost associated with LBP exceeds US $100 billion per year [5,6]. It is also a significant contributor to the current global epidemic of narcotic prescriptions [7].

Optimizing management of conditions such as LBP requires consumers to be engaged in their care. To enable this, health care providers need to have an understanding of the full context of the condition from the consumer perspective. “Contextual variables” here refer to any type of useful information about the context of an individual’s pain experience, such as physical, emotional, social, and/or occupational variables [8]. A better understanding of the contextual variables of individuals with LBP could provide clinicians and health providers with an alternative insight into patients’ concerns, beliefs, and expectations, and has the potential to improve outcomes in LBP [9]. Although there have been many studies examining individuals’ beliefs about LBP, patients’ perspectives remain inadequately understood [10]. Although qualitative studies—including systematic scoping reviews—investigating patients' needs and expectations have been conducted, these have largely focused on a single topic, such as health care, with the findings extrapolated from heterogeneous studies that are of poor quality [11-13]. A further limitation of current approaches is that most traditional data collection methods use predefined frameworks that have the potential to constrain responses. For instance, validated questionnaires that provide statements about back pain and its consequences (such as “back pain must be rested”) and require the respondent to indicate their level of agreement on a scale are commonly used [12,13]. Moreover, for logistical and methodological reasons, many studies restrict the selection of populations to be studied.

With the current advances in online and web technologies, social media has emerged as a new and rich source of first-person health care data [14-16]. Social media platforms provide an opportunity to rapidly collect data from a larger and more diverse population in a cost-efficient manner. Health-related topics are commonly discussed on Twitter [17-19], a microblogging social media site [20]. A systematic review conducted by Sinnenberg et al [21] found six main uses of Twitter in health research: content analysis, surveillance, engagement, recruitment, intervention, and network analysis. Aggregation and analysis of large volumes of health-related data from social media sites could provide valuable information from a first-person point of view [14,22]. In the area of LBP, this approach could be used to investigate individuals’ perspectives and the context around the LBP experience [15,23]. We hypothesize that the detected topics identify specific contexts around the LBP experience in individuals. Thus, the aim of this study was to identify contextual variables of the LBP experience from a first-person perspective using a topic modeling approach of Twitter data to provide useful insights into individuals’ beliefs and perceptions. This has the potential to inform more effective patient-centered approaches to the management of LBP.

Methods

Study Approach

Our study approach was to undertake content analysis of Twitter data by applying topic modeling. Content analysis is a widely used technique for qualitative research [24] that enables studying patient experience in depth by deriving topics of interest from text documents [14,25].

Twitter Data

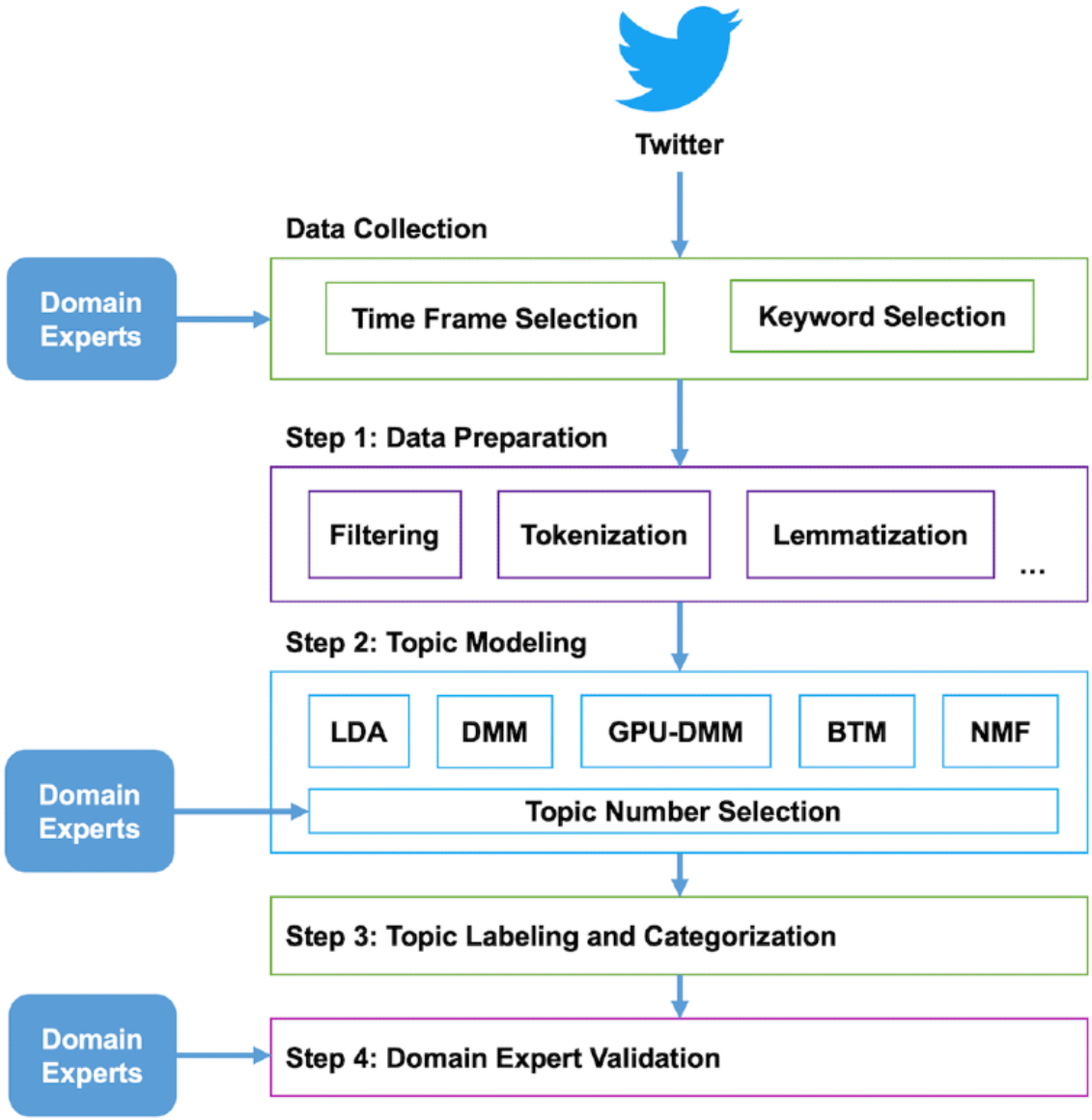
Twitter was used as the data source rather than other social media platforms, blog posts, or news articles because individuals use this platform for expressing and sharing their feelings and opinions on health-related topics by posting short messages that can be easily collected through application programming interfaces (APIs) or other open sources [14-17,26]. We used an open-source Twitter scraping tool called Twint [27] for collecting tweets related to LBP that were written in English. Twint enables the collecting of Twitter data without using Twitter's API through its publicly available library in the Python programming language [27,28]. We collected tweets posted between January 1, 2014, and December 31, 2018 (inclusive). The time frame of 5 years was selected to provide us with sufficient data to examine the patterns in emerging topics and the number of tweets over time. Since the number of active users on the social media platform increased in recent years and we needed a large volume of data for topic modeling, we did not consider tweets posted before 2014. We selected the search keywords based on 3 studies on back pain [15,29,30]. These are detailed in Table 1. Search keywords were verified by our domain experts (FC, a rheumatologist; DU, a physiotherapist) who have extensive research and clinical expertise in the area of LBP. Selecting search keywords and an appropriate time frame are important considerations in the data collection process. The Monash University Human Research Ethics Committee approved this study (project ID 19738).

Our data processing and analysis consisted of 4 steps (see Figure 1).

Table 1. Keywords used to search tweets related to low back pain.

Source	Study purpose	Keywords	Total, n
Lee et al, 2016 [15]	To quantify the risks associated with a new tweet about back pain	"painful back," "sore back," "back started hurting," "buggered my back," "hurt my back," "I've got backache," "injured my back," "my back hurts," "I've got back pain," "pain in my back," "put my back out," "my back is killing me"	12
Ahlwardt et al, 2014 [30]	To compare self-reported toothache experiences in tweets with those of backache, earache, and headache	"backache," "back ache," "back aches," "back hurt," "back hurting," "back hurts," "back killin'," "back killing," "back pain," "back sore"	10
Campbell et al, 2013 [29]	A systematic review to study the influence of employment social support in nonspecific back pain	"lumbago," "backache," "back ache," "back pain," "low back ache," "low back pain," "lower back pains"	7

Figure 1. The overall data analysis workflow. The analysis consists of four steps: (1) data preprocessing, (2) thematic analysis using topic modelling, (3) topic labeling and categorization, and (4) domain expert validation. BTM: biterm topic model; DMM: Dirichlet multinomial mixture; GPU-DMM: General Pólya Urn Dirichlet Multinomial Mixture; LDA: latent Dirichlet allocation; NMF: nonnegative matrix factorization.



Step 1: Data Preprocessing

We removed duplicates, retweets, URLs, and tweets related to marketing and advertisements, which reduced the data set from 7,892,210 to 2,825,645. We filtered the data further by removing tweets that did not contain first person pronouns [15]. As a result, the remaining data set size was 2,010,295.

We replaced contractions with their expanded forms (eg, "didn't" to "did not"). We converted the HTML characters to ASCII characters and removed hashtags, Unicode strings (eg, "\u2026"), numbers, and punctuation. We replaced abbreviations, elongated words (eg, "gooood" to "good"), and emoticons and emojis with their equivalent English expressions. We then performed spelling correction, lowercasing, tokenization, and lemmatization, created n-grams, removed stop words (eg, common terms such as "the" and "is"). We again removed the duplicates and the remaining data set was 1,249,576 tweets.

After completing the abovementioned steps, we excluded tweets with less than three words because in topic modeling, the document size is important to achieve high accuracy [31]. This reduced the data set to 896,867 tweets.

Step 2: Topic Modeling

Topic modeling is a technique used to provide a summary of a large collection of documents by extracting "topics" that represent the dominant themes [32]. It allows the uncovering of common, hidden themes from a corpus of text documents such as tweets. We tested 5 well-established topic modeling algorithms for detecting topics in a text-based corpus, namely latent Dirichlet allocation (LDA) [33], Dirichlet multinomial mixture (DMM) [34], GPU-DMM [35], biterm topic model (BTM) [36], and nonnegative matrix factorization (NMF) [37].

LDA is a generative probabilistic model that assumes each document can be represented by distribution over topics and each topic by distribution over words [33,38]. DMM is also a generative model but it assumes that each document is associated with one single topic [34,39]. GPU-DMM is an extended method of DMM that considers semantic similarity between words to provide semantic understanding of text documents and improve topic inference [35,40]. BTM uncovers topics by modeling the word co-occurrence patterns (ie, biterms) rather than using the document-level word co-occurrences [36,41]. NMF is able to learn the latent features in data using a nonnegative representation and improve latent semantic topic identification [37,42,43].

To use these models (except for NMF), we used a Java-based open-source library for short text topic modeling algorithms called STTM (version 1.8) [44], whereas for NMF we used the sklearn [45] library. For each approach, we performed a series of experiments ranging from 5 topics to 200 topics. We applied the 5 algorithms to the 896,867 tweets to determine the best model and the optimal number of topics.

Choosing the right number of topics is a crucial step in topic modeling because it can affect the accuracy of results. The quantitative approach computes the coherence score and perplexity, which helps in determining the optimal number of topics [46]. The coherence score measures the sum of the pairwise word-similarity scores of the words in the topic, using the pointwise mutual information (PMI) score [47]. Best collocation pairs usually have a high PMI. On the other hand, the qualitative approach requires humans and domain experts to examine the topics. Human judgment is extremely important because topic modeling uses a form of unsupervised learning.

As a quantitative approach, we calculated the coherence score of each model on different numbers of topics ranging from 5 to 200, based on the PMI score [47,48]. The coherence score was used to evaluate the quality of the topic-word distribution. LDA outperformed other approaches (ie, DMM, GPU-DMM, BTM, and NMF).

Additionally, we used a qualitative approach to select the most representative topics. We manually examined the topics, their top 20 terms, and a random sample of tweets in each topic. We also created a word cloud for each topic and evaluated word clouds and their sample tweets. We identified the number of topics that provided us with distinct and meaningful topics; if we exceeded this number of topics, we started to notice an increase in duplicates and overlapping topics. We used both quantitative and qualitative approaches to select the optimal number of topics.

Step 3: Topic Labeling and Categorization

Topic labeling is a process of representing the meaning of a topic by assigning each topic a descriptive word or phrase [49]. Although automatic labeling approaches can reduce costs and time required, they are not able to achieve high semantic validity and accuracy [50,51]. In our study, we used the "eyeballing" method, which refers to reading and inspecting the top words in a topic and manually assigning a label [50]. We made sure that the results met the requirements of a "good" label: (1) semantically relevant, (2) meaningful, (3) representative, (4) adequate, and (5) understandable [34,49].

LDA assumes that each document (tweet) is a mix of topics with different proportions [33]. We were interested to examine tweets based on their dominant topic to gain a better understanding of the frequency of topics across all tweets. Therefore, we performed further analysis, and used the label of the dominant topic to represent each tweet, and then calculated the total number of tweets per topic.

To improve the results of thematic analysis, low-order topics can be grouped under broad, higher-order categories [52]. Higher-level categories can provide a better overview of the key topics discussed by individuals. To this end, after manual topic labeling, we performed topic categorization and assigned a category label to the topics that represented common themes. To identify the important and widely discussed categories, we then calculated the percentage of all tweets that corresponded to each individual category.

Step 4: Domain Expert Validation

Two domain experts (FC, a rheumatologist; DU, a physiotherapist), actively working clinically and researchers in the area of LBP, independently examined the selected topics from the previous step where each topic included the top 20

words to determine face validity. As previously described, in topic modeling, the top words of each topic provide the description of that topic, thereby assisting the domain experts with inferring its meaning [49]. The experts then met to reconcile any differences and develop the final labels.

Results

Overview

The total number of collected tweets about LBP was 7,892,210 from 2,420,258 unique users from 2014 to 2018. The average number of words in each tweet increased from 2017 onward (Multimedia Appendix 1), in line with Twitter doubling the character limit of tweets from 140 to 280 characters as of November 2017 [53].

Step 1: Data Preprocessing

After performing comprehensive data preprocessing, the final number of retained tweets was 896,867, which represents 11% (896,867/7,892,210) of the original raw data we collected, with a vocabulary size of 29,539. The minimum length of tweets was 4 words and the maximum length was 20 words.

Step 2: Topic Modeling

After testing 5 topic modeling algorithms and the number of topics based on the coherence score and our manual examination, we selected the best model that included 60 topics, detected from 896,867 self-reported tweets about LBP. Multimedia Appendix 2 shows the coherence score of different models with a different number of topics ranging from 5 to 200. The best model was the LDA model with 60 topics, which had a coherence score of 0.562. Multimedia Appendix 3 shows the best model selected with 60 topics and their top 20 terms.

Step 3: Topic Labeling and Categorization

The 60 topics were examined and manually given a topic label. The common and duplicate labels were then grouped into higher-order categories. Word clouds for the two categories of “pain regions” and “sleep” after combining the related topics are provided in Multimedia Appendix 4. The prevalence of the 60 manually labeled topics is presented in Multimedia Appendix 5.

Step 4: Domain Expert Validation

Independent examination of selected topics by two domain experts and reconciliation of any differences resulted in 19 contextual categories, with details presented in Multimedia Appendix 6. The total number of tweets within each of 19 contextual categories is presented in Figure 2, with more details in Multimedia Appendix 7. The “emotion and beliefs” category had the largest proportion of the total tweets, followed by “physical activity” and “daily life.” The lowest proportion of tweets belonged to the categories of “food and drink,” “weather,” and “not being understood.”

The proportion of tweets for each higher-level category over the years showed that all 19 categories had been discussed by individuals with relatively similar frequency every year (see Figure 3). However, the proportion of “emotion and beliefs” decreased from 2014 to 2018. The number of tweets about other categories, such as “aggravating factors” and “symptoms,” increased over that time period. An example of a tweet for each category is presented in Table 2 to illustrate the type of personal point of view related to each category.

Figure 2. The 19 categories and their proportions based on all tweets posted from 2014 to 2018.

Figure 3. The proportions of 19 categories based on the dominant topic per year.

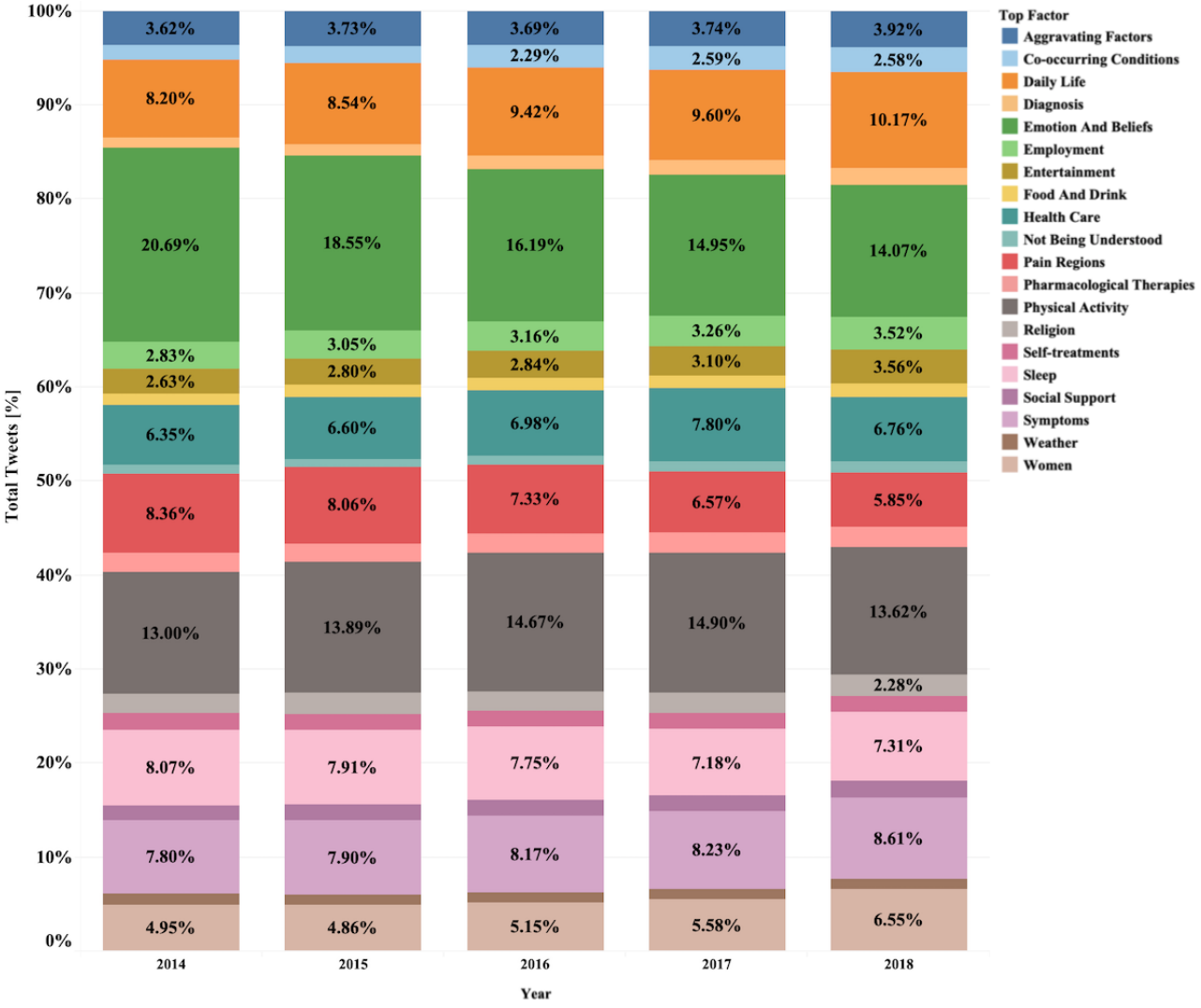


Table 2. An example of tweets for each contextual category.

Categories	Examples of tweets
Emotion and beliefs	• My back hurts, feeling sad because I wanna get up and do something ! I hate staying in bed :(
Physical activity	• I did 6 miles on my exercise bike yesterday, felt really pleased with myself, and ate healthy. My back hurts today
Daily life symptoms	• So my back hurts like hell and I can hardly sit here and do my hair. • I hate it when my lower back hurts and sends shooting pains down my legs, making them ache and throb. Ugh.
Sleep	• Every time I sleep in my sis guest bedroom my back hurts, that bed is not comfortable. I"d prolly be better off sleeping on the floor
Pain regions	• today is not a good day. my back hurts, my shoulder hurts, my elbow is tingly, a little numb down to my hand and to top it off now my left knee hurts a little.
Health care	• So I have found one good physio and one good chiropracter, both same price, who would you see if you had lower back pain?
Women	• Being pregnant is literally taking everything out of me. I'm exhausted, my back is killing me and I stay moody…
Aggravating factors	• Yesterday I tried doing a back flip on my trampoline. Now, every time I walk my back hurts. When I did the back flip I landed on my head.
Employment	• Hurt my back at work yesterday and I'm working a full 12 hours tomorrow without getting paid. Lovin life right now.
Entertainment	• Watching Cirque Du Soleil: Michael Jackson my back hurts just from watching it
Religion	• Testimony Time! i want to give God the glory for healing me from a severe back pain
Co-occurring conditions	• I don't know if my back pain is causing depression or my depression is causing back pain…
Pharmacological therapies	• I just took my very first Oxycodone for lower back pain. I think I'm in love. It didn't just kill the pain. It assassinated it.
Self-treatments	• Coconut oil epsom salt & vapor bath oil just soothed my back pain away
Social support	• Told mom my back hurts she offered to rub my feet an back I have the best mom ever
Food and drink	• my back is killing me cant get out ov bed but need coffee
Weather	• I love cold weather but it's really not helping with my back pain. Where is that warm summer weather atttttttt.
Not being understood	• OMG no one understands the pain I'm in right now. My back is killing me.

Discussion

Principal Results

In this study, we identified 60 specific topics from 896,867 tweets about LBP and grouped them into 19 categories that relate to contextual variables of LBP. The top category was "emotion and beliefs," with 157,563/896,867 tweets (17.6%), followed by "physical activity" (124,251/896,867, 13.85%) and "daily life" (80,730/896,867, 9%), while "food and drink," "weather," and "not being understood" had the lowest proportions of tweets (11,551/896,867, 1.29%; 10,109/896,867, 1.13%; and 9180/896,867, 1.02%, respectively). There were 11 topics within the category of "emotion and beliefs"; of 157,563 tweets in this category, 113,562 (72%) expressed negative sentiment. Our results were consistent with the general findings from traditional study methods in the area of LBP but provided more in-depth detail on the context of LBP from the individual perspective.

Comparison With Prior Work

Our study examined contextual variables to provide a novel insight into first-person perspectives of the LBP experience and confirmed the broad areas that have previously been identified using more traditional data collection methods from qualitative and quantitative studies. For example, psychosocial factors have an important role in LBP [54] and, from our analysis of tweets, "emotion and beliefs" was the most common topic we identified, with 157,563 of 896,867 tweets. This is consistent with LBP being widely recognized as a biopsychosocial condition, and growing evidence to show that psychological factors, such as beliefs and emotions, play an important role [55]. For instance, systematic reviews have highlighted that beliefs about back

pain and negative consequences resulting from these beliefs are common across different countries and populations [56], and affect both treatment efficacy and prognosis [57]. Moreover, mass media campaigns that target negative beliefs have been implemented in an effort to influence how people manage their back pain on a population level [58]. Our study has also provided novel findings with respect to emotions. Although we found a range of emotions, from positive emotions (such as happy, love, or fun) to negative emotions (including hell, bad, or disgusting), the majority were found to be of negative affect. Although several studies have examined the role of specific emotions, such as anger [59,60], in LBP research, our understanding of the array of emotions experienced by individuals with back pain, specifically negative emotions, is limited.

Our study also highlighted areas related to the pain experience in individuals that have not been adequately explored in the literature but that play an important role in the effectiveness of LBP interventions and self-management behaviors, such as the "not being understood," "religion," and "food and drink" categories. We found that although the category of "not being understood" had the smallest proportion of tweets with a total of 9180 tweets, it had the top five words: "make," "people," "stop," "thing," and "complain." This is consistent with a previous systematic scoping review that examined what patients want from their medical care, which reported that patients felt misunderstood and wanted legitimation of their LBP [11]. Patients with LBP report negative social stereotyping from health care professionals, family and friends, and the community [61] and that they are dissatisfied with the inadequate advice they receive from medical practitioners and have identified an unmet need for care providers that show more understanding and empathy [11].

The category of "food and drink" is novel and interesting. The tweets included words relating to the type of food (eg, pizza, chocolate, cookies and cream), mealtimes (such as breakfast and lunch), and the process of bringing or making food. Although they reflect important daily habits of eating and drinking, they may also highlight issues around pain affecting an individual's capacity to eat and drink and/or problems associated with weight and in particular obesity [62], which is a major public health issue [63].

There are well-described sex differences in the prevalence of back pain [64]. Analysis of tweets identified 3 topics under the "women" category including "motherhood," "large breasts problem," and "female health complaints." LBP has been reported in more than two-thirds of pregnant women [65]. Improving psychological well-being, physical fitness, and general well-being may reduce LBP in women [65-67]. The topics identified in tweets may provide more direction in relation to the personal topics that warrant further examination (eg, the potential effect of "large breasts problem" and whether this is a cause of LBP or a potential confounding variable). Identifying possible mechanisms for the association with topics such as "motherhood" or "female health complaints" could also help with understanding whether these associations are due to psychosocial factors or biomechanical factors such as the lifting and carrying of children. Understanding the context of LBP could offer valuable insights into how people with LBP view and experience their condition; this could lead to the identification of new areas of research in exploring the causes of LBP, as well as the opportunity to identify areas of potential misinformation that need to be addressed.

Limitations

There are some limitations to our study. Although the keywords were taken from existing studies about LBP and approved by domain experts, some keywords, such as "back hurt" and "back pain," were very broad. Therefore, the data collected might not have been specific to LBP. Selection of the right keywords in Twitter data analysis is very important to avoid unrelated data that could reduce the accuracy of results. Filtering and cleaning of Twitter data is also crucial for achieving high accuracy of results. In our study, we performed vigorous data cleaning, but our manual examination showed that there was a group of tweets that contained a few lines from the lyrics of a famous hip-hop song (Bad and Boujee) by Migos. These lines included "…So my money makin' my back ache." One of our search keywords was "back ache." Although there are many tools and methods available to automatically perform data cleaning, it is always necessary to manually inspect the results.

Twitter users tend to be younger and might not represent the general population; therefore, the results must be carefully interpreted [68]. Similar to other social media studies in health care, we cannot verify that individuals who tweeted about LBP were actually real patients [15]. However, the filtering based on first-person pronouns (eg, I, my, or mine) that we performed is likely to have reduced this.

To determine the optimal number of topics, we used the coherence score, a widely used method, and then manually examined and compared the models. This process can be further improved by using other measures such as heuristic approaches [69] or perplexity measures [70].

We also recognize that manual labeling of topics can be subjective. Two domain experts with extensive knowledge were involved in the labeling and examination of selected topics but future work in this area could involve a greater number of and more diverse domain experts to further reduce this subjectivity.

Conclusions

Our findings provided useful insights into individuals' beliefs and perspectives regarding their needs and concerns related to LBP that complement the information available in the literature. Considering the contextual factors identified in this study rather than simply focusing on a biomedical model of LBP could address the needs of patients more holistically, help with improving LBP outcomes, and increase patient satisfaction. These findings have the potential to assist health care providers and clinicians with developing more effective, personalized therapies for LBP. There is also the potential to use social media to identify any major changes in community beliefs and needs regarding LBP that can be addressed in a timelier manner.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. DU was supported by a National Health and Medical Research Council Career Development Fellowship (Level 2; 1142809).

Authors' Contributions

PDH, FB, DU, and FC contributed to study concept and design. R contributed to data collection and topic modeling. PDH, DU, and FC contributed to topic labeling and clustering. PDH, FB, DU, and FC contributed to interpretation of data. R and PDH contributed to drafting of the initial manuscript. PDH, FB, DU, and FC contributed to critical revision of the manuscript for important intellectual content. R and PDH provided administrative, technical, or material support. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The average number of words in tweets per year.
[DOCX File , 83 KB - jmir_v23i12e26093_app1.docx]

Multimedia Appendix 2
Coherence score for latent Dirichlet allocation, Dirichlet multinomial mixture (DMM), General Pólya Urn Dirichlet Multinomial Mixture (GPU-DMM), biterm topic model, and nonnegative matrix factorization with number of topics 5-200 .
[DOCX File , 197 KB - jmir_v23i12e26093_app2.docx]

Multimedia Appendix 3
The best model selected with 60 topics and their top 20 terms.
[DOCX File , 35 KB - jmir_v23i12e26093_app3.docx]

Multimedia Appendix 4
Word clouds for the pain region and sleep categories.
[DOCX File , 244 KB - jmir_v23i12e26093_app4.docx]

Multimedia Appendix 5
Total number of tweets per each topic manually labelled.
[DOCX File , 23 KB - jmir_v23i12e26093_app5.docx]

Multimedia Appendix 6
The 19 contextual categories related to low back pain.
[DOCX File , 23 KB - jmir_v23i12e26093_app6.docx]

Multimedia Appendix 7
The total and percentage of tweets for each contextual category.
[DOCX File , 21 KB - jmir_v23i12e26093_app7.docx]

References

1. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jun 24;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428] [Medline: 24665116]
2. Global Burden of Disease Study 2013 Collaborators. 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 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
3. Rubin DI. Epidemiology and risk factors for spine pain. Neurol Clin 2007 May;25(2):353-371. [doi: 10.1016/j.ncl.2007.01.004] [Medline: 17445733]
4. Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. Best Pract Res Clin Rheumatol 2010 Dec;24(6):769-781. [doi: 10.1016/j.berh.2010.10.002] [Medline: 21665125]

5. Guo HR, Tanaka S, Halperin WE, Cameron LL. Back pain prevalence in US industry and estimates of lost workdays. Am J Public Health 1999 Jul;89(7):1029-1035. [doi: 10.2105/ajph.89.7.1029] [Medline: 10394311]
6. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. J Bone Joint Surg Am 2006 Apr;88 Suppl 2:21-24. [doi: 10.2106/JBJS.E.01273] [Medline: 16595438]
7. Mafi JN, McCarthy EP, Davis RB, Landon BE. Worsening trends in the management and treatment of back pain. JAMA Intern Med 2013 Sep 23;173(17):1573-1581 [FREE Full text] [doi: 10.1001/jamainternmed.2013.8992] [Medline: 23896698]
8. Chou L, Cicuttini FM, Urquhart DM, Anthony SN, Sullivan K, Seneviwickrama M, et al. People with low back pain perceive needs for non-biomedical services in workplace, financial, social and household domains: a systematic review. J Physiother 2018 Apr;64(2):74-83 [FREE Full text] [doi: 10.1016/j.jphys.2018.02.011] [Medline: 29574167]
9. Lim YZ, Chou L, Au RT, Seneviwickrama KMD, Cicuttini FM, Briggs AM, et al. People with low back pain want clear, consistent and personalised information on prognosis, treatment options and self-management strategies: a systematic review. J Physiother 2019 Jul;65(3):124-135 [FREE Full text] [doi: 10.1016/j.jphys.2019.05.010] [Medline: 31227280]
10. De Souza LH, Frank AO. Subjective pain experience of people with chronic back pain. Physiother Res Int 2000 Nov;5(4):207-219. [doi: 10.1002/pri.201] [Medline: 11129663]
11. Chou L, Ranger TA, Peiris W, Cicuttini FM, Urquhart DM, Sullivan K, et al. Patients' perceived needs of health care providers for low back pain management: a systematic scoping review. Spine J 2018 Apr;18(4):691-711. [doi: 10.1016/j.spinee.2018.01.006] [Medline: 29373836]
12. Symonds TL, Burton AK, Tillotson KM, Main CJ. Absence resulting from low back trouble can be reduced by psychosocial intervention at the work place. Spine (Phila Pa 1976) 1995 Dec 15;20(24):2738-2745. [doi: 10.1097/00007632-199512150-00016] [Medline: 8747253]
13. Waddell G, Newton M, Henderson I, Somerville D, Main C. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993 Feb;52(2):157-168. [doi: 10.1016/0304-3959(93)90127-B] [Medline: 8455963]
14. Delir Haghighi P, Kang Y, Buchbinder R, Burstein F, Whittle S. Investigating Subjective Experience and the Influence of Weather Among Individuals With Fibromyalgia: A Content Analysis of Twitter. JMIR Public Health Surveill 2017 Jan 19;3(1):e4 [FREE Full text] [doi: 10.2196/publichealth.6344] [Medline: 28104577]
15. Lee H, McAuley JH, Hübscher M, Allen HG, Kamper SJ, Moseley GL. Tweeting back: predicting new cases of back pain with mass social media data. J Am Med Inform Assoc 2016 May;23(3):644-648. [doi: 10.1093/jamia/ocv168] [Medline: 26661720]
16. Asghar MZ, Ahmad S, Qasim M, Zahra SR, Kundi FM. SentiHealth: creating health-related sentiment lexicon using hybrid approach. Springerplus 2016;5(1):1139 [FREE Full text] [doi: 10.1186/s40064-016-2809-x] [Medline: 27504237]
17. 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]
18. Bian J, Topaloglu U, Yu F. Towards Large-scale Twitter Mining for Drug-related Adverse Events. SHB12 (2012) 2012 Oct 29;2012:25-32 [FREE Full text] [doi: 10.1145/2389707.2389713] [Medline: 28967001]
19. 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]
20. Aichner T, Jacob F. Measuring the Degree of Corporate Social Media Use. International Journal of Market Research 2015 Mar 01;57(2):257-276. [doi: 10.2501/ijmr-2015-018]
21. 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 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
22. Jayaraman PP, Forkan ARM, Morshed A, Haghighi PD, Kang Y. Healthcare 4.0: A review of frontiers in digital health. WIREs Data Mining Knowl Discov 2019 Dec 25;10(2):1-23. [doi: 10.1002/widm.1350]
23. Goh T, Delir Haghighi P, Burstein F, Buchbinder R. Developing a Contextual Model towards Understanding Low Back Pain. In: Proceedings of the 19th Pacific Asia Conference on Information Systems. 2015 Presented at: 19th Pacific Asia Conference on Information Systems; July 5, 2015; Singapore URL: http://aisel.aisnet.org/pacis2015/64
24. 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]
25. Hewis J. Do MRI Patients Tweet? Thematic Analysis of Patient Tweets About Their MRI Experience. J Med Imaging Radiat Sci 2015 Dec;46(4):396-402. [doi: 10.1016/j.jmir.2015.08.003] [Medline: 31052120]
26. Prier K, Smith M, Giraud-Carrier C, Hanson C. Identifying Health-Related Topics on Twitter: An Exploration of Tobacco-Related Tweets as a Test Topic. In: Salerno J, Yang SJ, Nau D, Chai SK, editors. Social Computing, Behavioral-Cultural Modeling and Prediction. Berlin, Heidelberg: Springer; Mar 19, 2011:18-25.
27. TWINT - Twitter Intelligence Tool. GitHub. URL: https://github.com/twintproject/twint [accessed 2021-03-29]
28. Xavier C, Souza M. A Basic Approach for Extracting and Analyzing Data from Twitter. In: Roesler V, Barrére E, Willrich R, editors. Special Topics in Multimedia, IoT and Web Technologies. Cham: Springer; Mar 3, 2020:185-211.
29. Campbell P, Wynne-Jones G, Muller S, Dunn KM. The influence of employment social support for risk and prognosis in nonspecific back pain: a systematic review and critical synthesis. Int Arch Occup Environ Health 2013 Feb 9;86(2):119-137 [FREE Full text] [doi: 10.1007/s00420-012-0804-2] [Medline: 22875173]

30. Ahlwardt K, Heaivilin N, Gibbs J, Page J, Gerbert B, Tsoh JY. Tweeting about pain: comparing self-reported toothache experiences with those of backaches, earaches and headaches. J Am Dent Assoc 2014 Jul;145(7):737-743 [FREE Full text] [doi: 10.14219/jada.2014.30] [Medline: 24982280]
31. Jian T, Zhaoshi M, Xuanlong N, Qiaozhu M, Ming Z. Understanding the limiting factors of topic modeling via posterior contraction analysis. 2014 Jun 21 Presented at: 31st International Conference on International Conference on Machine Learning; 2014; Beijing, China p. 190-198.
32. Blei D, Carin L, Dunson D. Probabilistic Topic Models: A focus on graphical model design and applications to document and image analysis. IEEE Signal Process Mag 2010 Nov 01;27(6):55-65 [FREE Full text] [doi: 10.1109/MSP.2010.938079] [Medline: 25104898]
33. Blei D, Ng A, Jordan M. Latent dirichllocation. Journal of Machine Learning Research 2003;3:993-1022 [FREE Full text]
34. Nigam K, Mccallum AK, Thrun S, Mitchell T. Text Classification from Labeled and Unlabeled Documents using EM. Machine Learning 2000;39:103-134. [doi: 10.21236/ada350490]
35. Li C, Wang H, Zhang Z, Sun A, Ma Z. Topic Modeling for Short Texts with Auxiliary Word Embeddings. In: SIGIR '16: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. New York, NY: ACM Press; 2016 Jul 17 Presented at: SIGIR '16; July 17 - 21; Pisa, Italy p. 165-174. [doi: 10.1145/2911451.2911499]
36. Cheng X, Yan X, Lan Y, Guo J. BTM: Topic Modeling over Short Texts. IEEE Trans Knowl Data Eng 2014 Dec 1;26(12):2928-2941. [doi: 10.1109/tkde.2014.2313872]
37. Lee D, Seung S. Algorithms for Non-negative Matrix Factorization. : MIT Press; 2000 Jan 1 Presented at: 13th International Conference on Neural Information Processing; 2000; Denver, CO p. 535-541.
38. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
39. 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]
40. Liang W, Feng R, Liu X, Li Y, Zhang X. GLTM: A Global and Local Word Embedding-Based Topic Model for Short Texts. IEEE Access 2018;6:43612-43621. [doi: 10.1109/access.2018.2863260]
41. Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R. Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access. J Med Internet Res 2018 Dec 27;20(4):e10029 [FREE Full text] [doi: 10.2196/10029] [Medline: 29613851]
42. Odlum M, Yoon S, Broadwell P, Brewer R, Kuang D. How Twitter Can Support the HIV/AIDS Response to Achieve the 2030 Eradication Goal: In-Depth Thematic Analysis of World AIDS Day Tweets. JMIR Public Health Surveill 2018 Nov 22;4(4):e10262 [FREE Full text] [doi: 10.2196/10262] [Medline: 30467102]
43. Wang Y, Zhang Y. Nonnegative Matrix Factorization: A Comprehensive Review. IEEE Trans Knowl Data Eng 2013 Jun;25(6):1336-1353. [doi: 10.1109/tkde.2012.51]
44. STTM: A Library of Short Text Topic Modeling. GitHub. URL: https://github.com/qiang2100/STTM [accessed 2021-02-17]
45. scikit-learn: Machine Learning in Python. URL: https://scikit-learn.org/stable/ [accessed 2021-02-17]
46. Chang J, Gerrish S, Wang C, Boyd-graber J, Blei D. Reading tea leaves: how humans interpret topic models. 2009 Presented at: 22nd International Conference on Neural Information Processing Systems; December 7, 2009; Vancouver, British Columbia, Canada p. 288-296 URL: https://papers.nips.cc/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html
47. Church K, Hanks P. Word association norms, mutual information, and lexicography. Computational Linguistics 1990 Mar 1;16(1):76-83. [doi: 10.3115/981623.981633]
48. Newman D, Bonilla E, Buntine W. Improving topic coherence with regularized topic models. 2011 Presented at: 24th International Conference on Neural Information Processing Systems; December 12, 2011; Granada, Spain p. 496-504 URL: https://papers.nips.cc/paper/2011/hash/5ef698cd9fe650923ea331c15af3b160-Abstract.html
49. Allahyari M, Pouriyeh S, Kochut K, Reza H. A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling. ijacsa 2017;8(9):335-349. [doi: 10.14569/ijacsa.2017.080947]
50. Morstatter F, Liu H. In Search of Coherence and Consensus: Measuring the Interpretability of Statistical Topics. Journal of Machine Learning Research 2017;18(1):6177-6208.
51. Grimmer J, Stewart BM. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 2017 Jan 04;21(3):267-297. [doi: 10.1093/pan/mps028]
52. Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. International Journal of Qualitative Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
53. Boot AB, Tjong Kim Sang E, Dijkstra K, Zwaan RA. How character limit affects language usage in tweets. Palgrave Commun 2019 Jul 9;5(1):1-13. [doi: 10.1057/s41599-019-0280-3]

54. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. Spine (Phila Pa 1976) 2002 Mar 01;27(5):E109-E120. [doi: 10.1097/00007632-200203010-00017] [Medline: 11880847]
55. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
56. Morton L, de Bruin M, Krajewska M, Whibley D, Macfarlane G. Beliefs about back pain and pain management behaviours, and their associations in the general population: A systematic review. Eur J Pain 2019 Jan 07;23(1):15-30 [FREE Full text] [doi: 10.1002/ejp.1285] [Medline: 29984553]
57. Wertli MM, Rasmussen-Barr E, Held U, Weiser S, Bachmann LM, Brunner F. Fear-avoidance beliefs-a moderator of treatment efficacy in patients with low back pain: a systematic review. Spine J 2014 Nov 01;14(11):2658-2678. [doi: 10.1016/j.spinee.2014.02.033] [Medline: 24614254]
58. Urquhart DM, Bell RJ, Cicuttini FM, Cui J, Forbes A, Davis SR. Negative beliefs about low back pain are associated with high pain intensity and high level disability in community-based women. BMC Musculoskelet Disord 2008 Nov 04;9(1):148-148 [FREE Full text] [doi: 10.1186/1471-2474-9-148] [Medline: 18983681]
59. Bruehl S, Liu X, Burns J, Chont M, Jamison R. Associations between daily chronic pain intensity, daily anger expression, and trait anger expressiveness: an ecological momentary assessment study. Pain 2012 Dec;153(12):2352-2358 [FREE Full text] [doi: 10.1016/j.pain.2012.08.001] [Medline: 22940462]
60. Burns JW, Quartana P, Bruehl S. Anger suppression and subsequent pain behaviors among chronic low back pain patients: moderating effects of anger regulation style. Ann Behav Med 2011 Aug 5;42(1):42-54 [FREE Full text] [doi: 10.1007/s12160-011-9270-4] [Medline: 21544702]
61. Slade SC, Molloy E, Keating JL. Stigma experienced by people with nonspecific chronic low back pain: a qualitative study. Pain Med 2009 Jan 01;10(1):143-154. [doi: 10.1111/j.1526-4637.2008.00540.x] [Medline: 19222775]
62. Chou L, Brady S, Urquhart D, Teichtahl AJ, Cicuttini FM, Pasco JA, et al. The Association Between Obesity and Low Back Pain and Disability Is Affected by Mood Disorders: A Population-Based, Cross-Sectional Study of Men. Medicine (Baltimore) 2016 Apr;95(15):e3367 [FREE Full text] [doi: 10.1097/MD.0000000000003367] [Medline: 27082599]
63. Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health. Int J Surg Oncol (N Y) 2017 Aug;2(7):e17 [FREE Full text] [doi: 10.1097/IJ9.0000000000000017] [Medline: 29177227]
64. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299-299 [FREE Full text] [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
65. Liddle S, Pennick V. Interventions for preventing and treating low-back and pelvic pain during pregnancy. Cochrane Database Syst Rev 2015 Sep 30(9):CD001139 [FREE Full text] [doi: 10.1002/14651858.CD001139.pub4] [Medline: 26422811]
66. Brady SR, Hussain SM, Brown WJ, Heritier S, Wang Y, Teede H, et al. Course and Contributors to Back Pain in Middle-aged Women Over 9 Years: Data From the Australian Longitudinal Study on Women's Health. Spine (Phila Pa 1976) 2018 Dec 01;43(23):1648-1656. [doi: 10.1097/BRS.0000000000002702] [Medline: 29794589]
67. Ng SK, Cicuttini FM, Davis SR, Bell R, Botlero R, Fitzgibbon BM, et al. Poor general health and lower levels of vitality are associated with persistent, high-intensity low back pain and disability in community-based women: A prospective cohort study. Maturitas 2018 Jul;113:7-12. [doi: 10.1016/j.maturitas.2018.04.007] [Medline: 29903650]
68. Zhang H, Wheldon C, Dunn A, Tao C, Huo J, Zhang R, et al. Mining Twitter to assess the determinants of health behavior toward GPU-DMM: General Pólya Urn Dirichlet Multinomial Mixturehuman papillomavirus vaccination in the United States. J Am Med Inform Assoc 2020 Feb 01;27(2):225-235 [FREE Full text] [doi: 10.1093/jamia/ocz191] [Medline: 31711186]
69. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16 Suppl 13:S8 [FREE Full text] [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]
70. Wallach H, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning.: ACM Press; 2009 Presented at: 26th Annual International Conference on Machine Learning; June 14-18, 2009; Montreal, Quebec, Canada p. 1-8.

Abbreviations

API: application programming interface
BTM: biterm topic model
DMM: Dirichlet multinomial mixture
GPU-DMM: General Pólya Urn Dirichlet Multinomial Mixture
LBP: low back pain
LDA: latent Dirichlet allocation
NMF: nonnegative matrix factorization

PMI: pointwise mutual information
STTM: short text topic modeling algorithm

Edited by JMIRPE Office; submitted 27.11.20; peer-reviewed by G Chirambo, E O'Hagan; comments to author 22.01.21; revised version received 06.03.21; accepted 21.11.21; published 23.12.21.

Please cite as:
Robert, Delir Haghighi P, Burstein F, Urquhart D, Cicuttini F
Investigating Individuals' Perceptions Regarding the Context Around the Low Back Pain Experience: Topic Modeling Analysis of Twitter Data
J Med Internet Res 2021;23(12):e26093
URL: https://www.jmir.org/2021/12/e26093
doi:10.2196/26093
PMID:36260398

Original Paper

COVID-19–Related Rumor Content, Transmission, and Clarification Strategies in China: Descriptive Study

Peishan Ning[1*], PhD; Peixia Cheng[1*], PhD; Jie Li[1*], BSc; Ming Zheng[1], BSc; David C Schwebel[2], PhD; Yang Yang[3], PhD; Peng Lu[4], PhD; Li Mengdi[4], MD; Zhuo Zhang[4], MD; Guoqing Hu[1,5], PhD

[1]Department of Epidemiology and Health Statistics, Hunan Provincial Key Laboratory of Clinical Epidemiology, Xiangya School of Public Health, Central South University, Changsha, China

[2]Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

[4]Department of Sociology, Central South University, Changsha, China

[5]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Guoqing Hu, PhD
Department of Epidemiology and Health Statistics
Hunan Provincial Key Laboratory of Clinical Epidemiology
Xiangya School of Public Health, Central South University
110 Xiangya Road
Changsha, 410078
China
Phone: 86 731 84805414
Email: huguoqing009@gmail.com

Abstract

Background: Given the permeation of social media throughout society, rumors spread faster than ever before, which significantly complicates government responses to public health emergencies such as the COVID-19 pandemic.

Objective: We aimed to examine the characteristics and propagation of rumors during the early months of the COVID-19 pandemic in China and evaluated the effectiveness of health authorities' release of correction announcements.

Methods: We retrieved rumors widely circulating on social media in China during the early stages of the COVID-19 pandemic and assessed the effectiveness of official government clarifications and popular science articles refuting those rumors.

Results: We show that the number of rumors related to the COVID-19 pandemic fluctuated widely in China between December 1, 2019 and April 15, 2020. Rumors mainly occurred in 3 provinces: Hubei, Zhejiang, and Guangxi. Personal social media accounts constituted the major source of media reports of the 4 most widely distributed rumors (the novel coronavirus can be prevented with "Shuanghuanglian": 7648/10,664, 71.7%; the novel coronavirus is the SARS coronavirus: 14,696/15,902, 92.4%; medical supplies intended for assisting Hubei were detained by the local government: 3911/3943, 99.2%; asymptomatically infected persons were regarded as diagnosed COVID-19 patients with symptoms in official counts: 322/323, 99.7%). The number of rumors circulating was positively associated with the severity of the COVID-19 epidemic (ρ=0.88, 95% CI 0.81-0.93). The release of correction articles was associated with a substantial decrease in the proportion of rumor reports compared to accurate reports. The proportions of negative sentiments appearing among comments by citizens in response to media articles disseminating rumors and disseminating correct information differ insignificantly (both correct reports: χ_1^2=0.315, *P*=.58; both rumors: χ_1^2=0.025, *P*=.88; first rumor and last correct report: χ_1^2=1.287, *P*=.26; first correct report and last rumor: χ_1^2=0.033, *P*=.86).

Conclusions: Our results highlight the importance and urgency of monitoring and correcting false or misleading reports on websites and personal social media accounts. The circulation of rumors can influence public health, and government bodies should establish guidelines to monitor and mitigate the negative impact of such rumors.

(J Med Internet Res 2021;23(12):e27339) doi:10.2196/27339

KEYWORDS

COVID-19; rumor; strategy; China; social media

Introduction

Accurate, timely, and publicly accessible information is critical to effectively control serious public health emergencies such as the COVID-19 pandemic [1]. Unfortunately, necessary information is sometimes unavailable to the public, and some news reports are misleading or inaccurate. The absence of proper information offers fertile ground for the emergence and propagation of rumors (also called *fake news*). If rumors appear convincing and are not effectively refuted, they can create serious consequences such as poor health-related decisions and distrust of public health agencies by the public.

Over the past few decades, rumor-debunking strategies such as fact checking, corrections, and retractions (eg, deleting social media posts) have been implemented as postevent responses to counteract the impact of rumors [2-4]. The effectiveness of these strategies has not been assessed systematically; previous research focuses on strategies to mitigate public belief of the rumors [5-7] but not on the role of strategies in reducing the proportion of rumor reports and alleviating negative sentiments among the public. Partly due to lack of solid research evidence, the World Health Organization (WHO) concluded that health-related rumors were poorly managed in nearly all major public health emergency events during the 21st century [1].

During the COVID-19 pandemic, rumors have emerged and propagated in nearly every country [8]. The rapid development and widespread use of the internet, social media, and mobile phone technology facilitate the emergence and propagation of rumors, making prevention and control of rumors more challenging than they were even a decade ago. In fact, widespread fake news rumors, such as “COVID-19 is just a version of seasonal influenza” and “Hydroxychloroquine is an effective drug to treat COVID-19,” significantly compromised COVID-19 prevention and control efforts and influenced social stability in highly populated countries such as the United States and Brazil [9].

China began facing large waves of pandemic-related rumors early in 2020. Between February and May 2020, the Chinese government actively released multiple clarification and correction announcements to reduce the influence of those rumors [10]. No empirical research has examined the characteristics of the COVID-19–related rumors or the effectiveness of the counteractive measures taken by Chinese health authorities. Knowledge of these characteristics would be helpful, both for countries across the globe that continue to face COVID-19–related rumors and public health challenges as well as to handle rumor patterns that emerge from future public health emergencies.

This study was designed to examine the characteristics and propagation of rumors during the early months of the COVID-19 pandemic in China and to evaluate the effectiveness of health authorities’ correction announcements. We considered the following research questions: What were the contents of rumors about COVID-19 in the early months of the pandemic in China, and what was the source of those rumors? Was the release of correction announcements by health authorities effective in mitigating the impact of major rumors?

Methods

Classification of the COVID-19 Rumors

Rumors were defined as reports that disseminate false information—information that is inconsistent with existing scientific evidence. In this study, all COVID-19 rumors were confirmed to be incorrect based on current scientific evidence by health authorities. In 3 rounds of discussions, we categorized the rumors that we identified into 8 groups: (1) prevention, (2) diagnosis/treatment/assistance, (3) origin and spread of COVID-19, (4) consequences of COVID-19, (5) disease statistics, (6) return to work or back to school, (7) imported cases, and (8) all others. We arranged a trained researcher to categorize each rumor into 1 of the 8 categories. One-third of the rumors were randomly extracted for a second evaluation by 3 other independent researchers. Consistency between both sets of evaluations was excellent (κ=0.96, P<.001).

Data Sources

Data sources were derived through 4 steps. First, based on preliminary search results, we identified 20 prominent and publicly accessible web-based platforms available in China to search for COVID-19–related rumors. These platforms included 9 websites, 5 official Sina Microblog accounts, and 6 WeChat public accounts (Table S1 in Multimedia Appendix 1).

Second, we retrieved all media reports regarding 4 major rumors and classified the media reports into 3 categories: those disseminating rumors, those providing correct information, and those with ambiguous information. We conducted this classification using the ZhongQing HuaYun web-based public service platform (CYYUN), a free public database platform licensed by the Central Committee of the Communist Youth League. CYYUN automatically retrieves media reports from WeChat accounts officially certified by governmental or business entities, Microblog accounts, public forums, news websites, print media, blog accounts, videos, news apps, major overseas media sources, and other mainstream national news sources. CYYUN data are updated approximately every 5 minutes.

Third, Sina Microblog Topic was used to gather readers’ comments in response to media reports of the rumor event with the most public comments. Sina Microblog Topic is a webpage that summarizes personal microblogs with a tag (“#topics-related keywords#”), allowing readers to post personal comments below the microblogs. Comment data from Sina Microblog Topic are freely accessible in China.

Last, the number of daily confirmed COVID-19 cases at the national and provincial levels were derived from the official website of the National Health Commission of China.

Data Collection

COVID-19 Rumor Reports

Given the pattern of the COVID-19 epidemic period in China, we limited the study time period from December 1, 2019 to April 15, 2020. We used Python (version 3.7) to develop a web crawler algorithm to automatically retrieve all media reports related to COVID-19 rumors from the 20 selected web-based platforms. Duplicate and irrelevant reports were removed by manually reviewing article titles and full-text articles.

Stages of the Early COVID-19 Epidemic in China

As defined by a previous study [11], the early phase of the COVID-19 pandemic in China was divided into 6 stages: (1) an early stage without any significant interventions (December 30, 2019-January 9, 2020); (2) massive population migration before the Spring Festival but no strong interventions were implemented (January 10, 2020-January 22, 2020); (3) city lockdowns, traffic suspension, and home quarantine (January 23, 2020-February 1, 2020); (4) centralized quarantine and treatment in designated hospitals or facilities, with improved medical resources (February 2, 2020-February 16, 2020); (5) centralized quarantine and whole-community symptom survey administered concerning COVID-19 symptoms, such as fever and respiratory symptoms (February 17, 2020-March 10, 2020); and (6) a focus on preventing imported cases (March 11, 2020-April 15, 2020).

Media Reports Related to Major COVID-19 Rumors

The number of relevant media reports obtained from CYYUN was used to estimate the social influence of rumors. We selected 1 rumor that was most socially influential from each of the 6 epidemic stages based on the number of relevant media reports. We retained 4 rumors for analysis; we excluded 2 rumors that attributed false quotes to Professor Nanshan Zhong, a famous Chinese scientist. The rationale of this exclusion is that the high social status of Professor Zhong rather than (or along with) the contents of these quotes might have fueled the spread of the 2 rumors, potentially leading to bias in our analysis of the impact of the contents of rumors.

We used CYYUN to collect all media reports related to the propagation of the 4 retained rumors: Case A, the novel coronavirus can be prevented with "Shuanghuanglian," a traditional Chinese medicine; Case B, the novel coronavirus is the SARS coronavirus; Case C, medical supplies intended for assisting Hubei were detained by the local government; and Case D, asymptomatically infected persons were regarded as diagnosed COVID-19 patients with symptoms in official counts. (Note that Case D is a rumor because the WHO and all major countries do not count asymptomatically infected persons as diagnosed COVID-19 cases in their official data counts [12]).

All media reports related to the 4 major rumors were gathered and processed. Based on official statements from the government regarding each rumor, trained researchers manually divided each media report into 3 categories: (1) the report disseminated or perpetuated the rumor, (2) the report disseminated correct information to refute the rumor, or (3) the report disseminated ambiguous information that did not clearly support or refute the rumor. To assure the consistency of the classifications, we randomly selected 500 media reports for each of the 4 major rumors and asked an independent researcher to categorize them into the 3 categories. Consistency between both sets of researchers was acceptable (Case A: κ=0.87, P<.001; Case B: κ=0.87, P<.001; Case C: κ=0.90, P<.001; Case D: κ=0.87, P<.001).

Microblog Readers' Comments

We used a Python crawler algorithm to retrieve reader comments concerning Microblog articles for the most discussed rumor (Case A) on Sina Microblog Topic, "The novel coronavirus can be prevented with "Shuanghuanglian." We selected Case A for this analysis because it was sustained for a longer time period than the other rumors. It also stimulated more media reports and readers' comments compared to the other 3 cases. Using the sentiment orientation analysis application programming interface from Baidu [13], we grouped reader comments into positive (expressing agreement, support, or optimistic attitudes, eg, "the situation will surely become better"); neutral (showing neither clear support and optimism, nor disagreement and pessimism, eg, "thanks for interacting with us"); or negative (expressing disagreement or worried, pessimistic, and ironic attitudes, eg, "I feel like swearing") categories.

Data Analysis

We implemented 2 strategies to assess the effectiveness of authorities' refutation of rumors.

We compared the distribution of the 2 types of media reports (those perpetuating the rumor and those correcting the rumor) before and after the rumor was formally refuted in the media. Because there is no standard criterion to determine the launch time of a formal rumor-refuting effort, we studied the temporal distributions of the number of correct reports and the number of rumor reports per hour for all 4 cases (Figure S1 in Multimedia Appendix 1). Based on these empirical distributions, we found that the number correct reports per hour generally increased substantially in subsequent hours once it reached over 50 reports. Thus, we considered the refutation as formally launched from the first hour in which more than 50 reports refuting the rumor were posted. We considered 3 phases for our analysis: prerefutation, refuting the rumor, and postrefutation.

We examined differences in the distribution of positive, neutral, or negative emotions in comments responding to reports disseminating the rumor and in reports correcting the rumor. We focused this analysis on case A, the rumor that received the most reader comments and was sustained the longest, thus having the largest social impact among all COVID-19 rumors during the early stages of the pandemic in China. To conduct the analysis, we examined both the frequency of reader comments and the time intervals between the first and last comments from the same readers when more than 2 comments were posted.

Based on publicly accessible Microblog users' unique identity accounts, we compared the distribution of comments in response to case A from readers who posted 2 comments or more. For each user, the first and last comments during the study period were considered, and the pairs were divided into 4 groups:

Group A, both comments occurred in response to reports disseminating correct information; Group B, both comments occurred in response to reports disseminating rumors; Group C, the first comment occurred in response to a report disseminating a rumor, and the last in response to correct information; and Group D, the first comment occurred in response to a report disseminating correct information and the last in response to a rumor. This grouping allowed us to examine the potential effect of the order of reading media reports (first reading reports disseminating rumors and then reading correct reports or vice versa).

All statistical analyses were performed using the R software package (version 4.0.0). Proportions and Spearman rank correlation coefficient with 95% confidence intervals were calculated. The chi-square test was used to test the difference in the proportion of rumor reports and correct reports between prerefutation and postrefutation. All tests were 2 tailed, and $P<.05$ was considered statistically significant.

Ethics Concerns

This study was approved by the ethics committee of Xiangya School of Public Health (XYGW-2020-43). This study used open-access social media data and excluded all personal information. The ethics committee determined that this study was exempt from requiring informed consent.

Results

Characteristic of Rumors

After screening 19,683 recorded rumors on 20 websites and social media accounts, we obtained 1829 unique COVID-19–related rumors in China that appeared from December 1, 2019 to April 15, 2020 (Figure S2 in Multimedia Appendix 1).

The frequency of rumors began with a quiescent period (<6 rumors per day before January 20, 2020), but then rose rapidly to reach an initial peak on January 25, 2020 (n=75). After a second peak on February 7, 2020 (n=82), the frequency started to decrease gradually (Figure 1a). “Pneumonia,” “WeChat,” and “spread” were the most commonly occurring words, with high-frequency words varying across the 6 stages of the COVID-19 epidemic (Table S2 and Figure S3 in Multimedia Appendix 1).

Figure 1b shows the large variation in the number of rumors across different provinces during the study period. Of the 1829 rumors, 399 (21.8%) rumors involved more than 1 province, but the majority (1430 rumors) involved a single province. The largest number occurred in Hubei (n=186), followed by Zhejiang (n=137) and Guangxi (n=121). Ningxia, Tibet, Xinjiang, Qinghai, Hainan, and Taiwan each had fewer than 10 rumors.

The content of the rumors evolved greatly across the 6 stages of the COVID-19 epidemic in China (Figure 1c). Rumors related to prevention and disease statistics were most common, accounting for 72.2% (39/54), 73.3% (335/457), and 64.8% (414/639) of all rumors in stages 2, 3, and 4, respectively. The percentage of rumors involving the categories *diagnosis/treatment/assistance* ($\chi_1^2=6.352$, $P=.01$), *return to work or back to school* ($\chi_1^2=148.094$, $P<.001$), and *imported cases* ($\chi_1^2=126.04$, $P<.001$) significantly increased during stages 5 and 6.

We considered in greater detail the 4 most widespread rumor cases (case A, the novel coronavirus can be prevented with “Shuanghuanglian”; case B, the novel coronavirus is the SARS coronavirus; case C, medical supplies intended for assisting Hubei were detained by the local government; and case D, asymptomatically infected persons were regarded as diagnosed COVID-19 patients with symptoms in official counts). Most reports were from personal social media accounts for case A (33,870/53,798, 63.0%), case B (21,234/24,436, 86.9%), and case C (6411/7101, 90.3%). For case D, most reports were from websites and social media accounts of news agencies (1850/3436, 53.8%) (Figure 2). Personal social media accounts were the major source of rumor reports (case A: 7648/10,664, 71.7%; case B: 14,696/15,902, 92.4%; case C: 3911/3943, 99.2%; case D: 322/323, 99.7%) and ambiguous reports (case A: 3116/4968, 62.7%; case B: 2582/2592, 99.6%; case C: 382/383, 99.7%; case D: 156/156, 100.0%).

In addition, there was a strong correlation between the frequency of rumors and the daily number of newly confirmed COVID-19 cases at the national level from December 30, 2019 to April 15, 2020 ($\rho=0.88$, 95% CI 0.81-0.93, $P<.001$). A spatial correlation between the frequency of rumors and the cumulative number of confirmed cases up to April 15, 2020 also emerged at the provincial level ($\rho=0.83$, 95% CI 0.61-0.94, $P<.001$) (Figure S4 in Multimedia Appendix 1).

Figure 1. Number of rumors related to the COVID-19 epidemic in China from December 30, 2019 to April 15, 2020 (a. time trend; b. provincial variations; c. contents by stage).

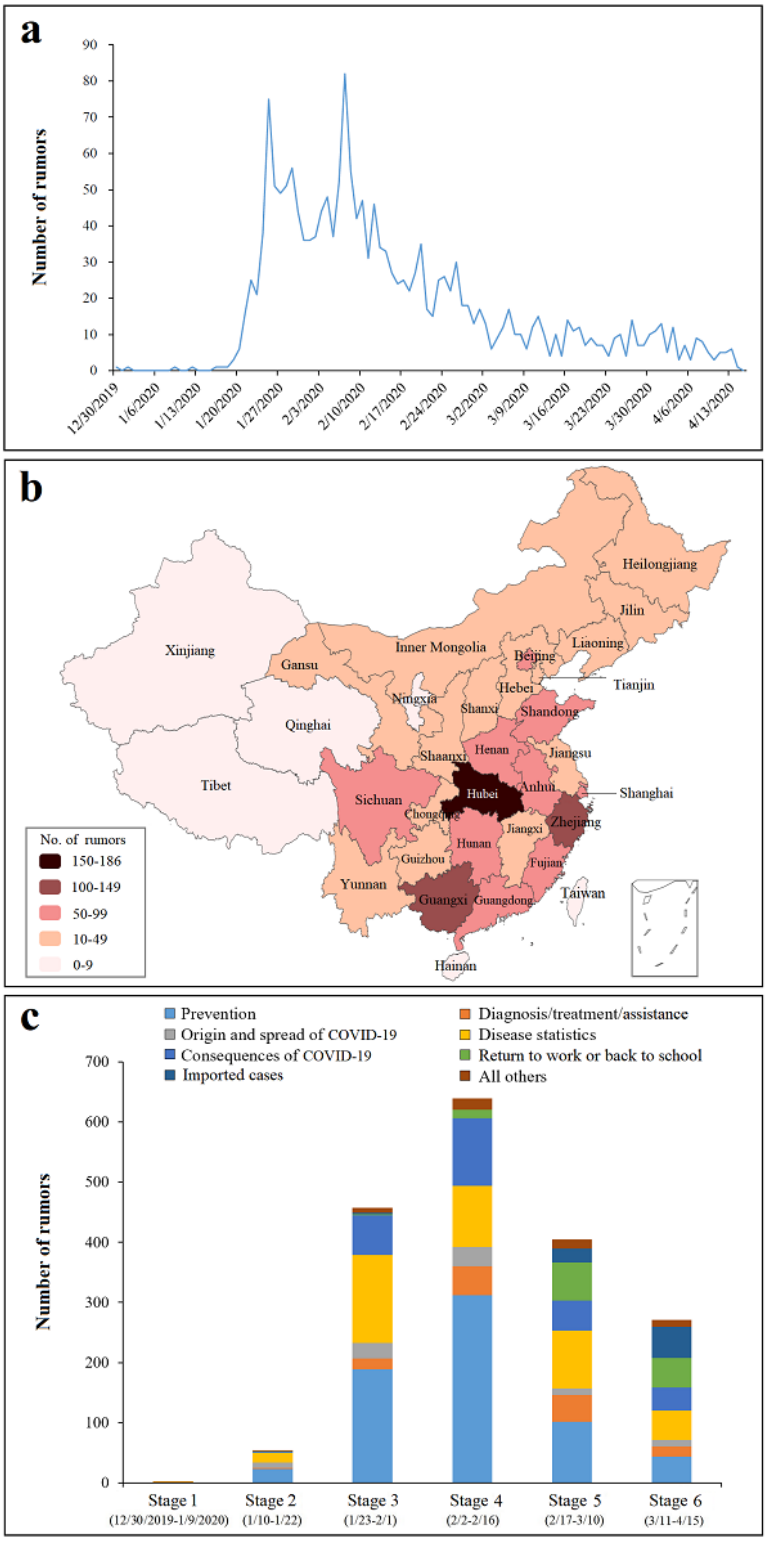
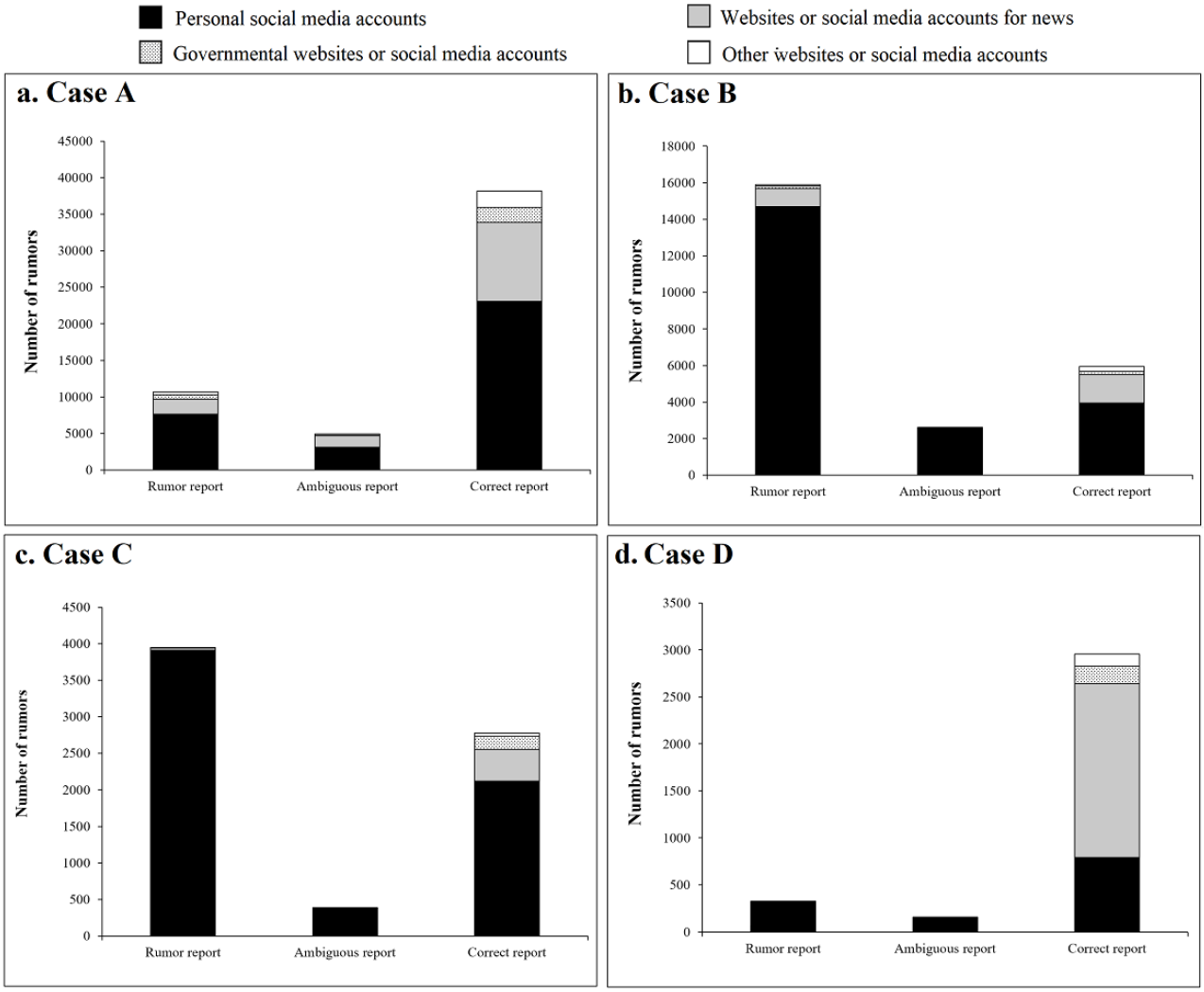


Figure 2. Source of media reports for 4 major rumors related to the COVID-19 epidemic in China.



Effectiveness of Releasing Correct Information

Totals of 53,798 media reports (case A), 24,436 media reports (case B), 7101 media reports (case C), and 3436 media reports (case D) were obtained after eliminating duplicates (n=3788) and irrelevant reports (n=4036) (Figure S2b in Multimedia Appendix 1).

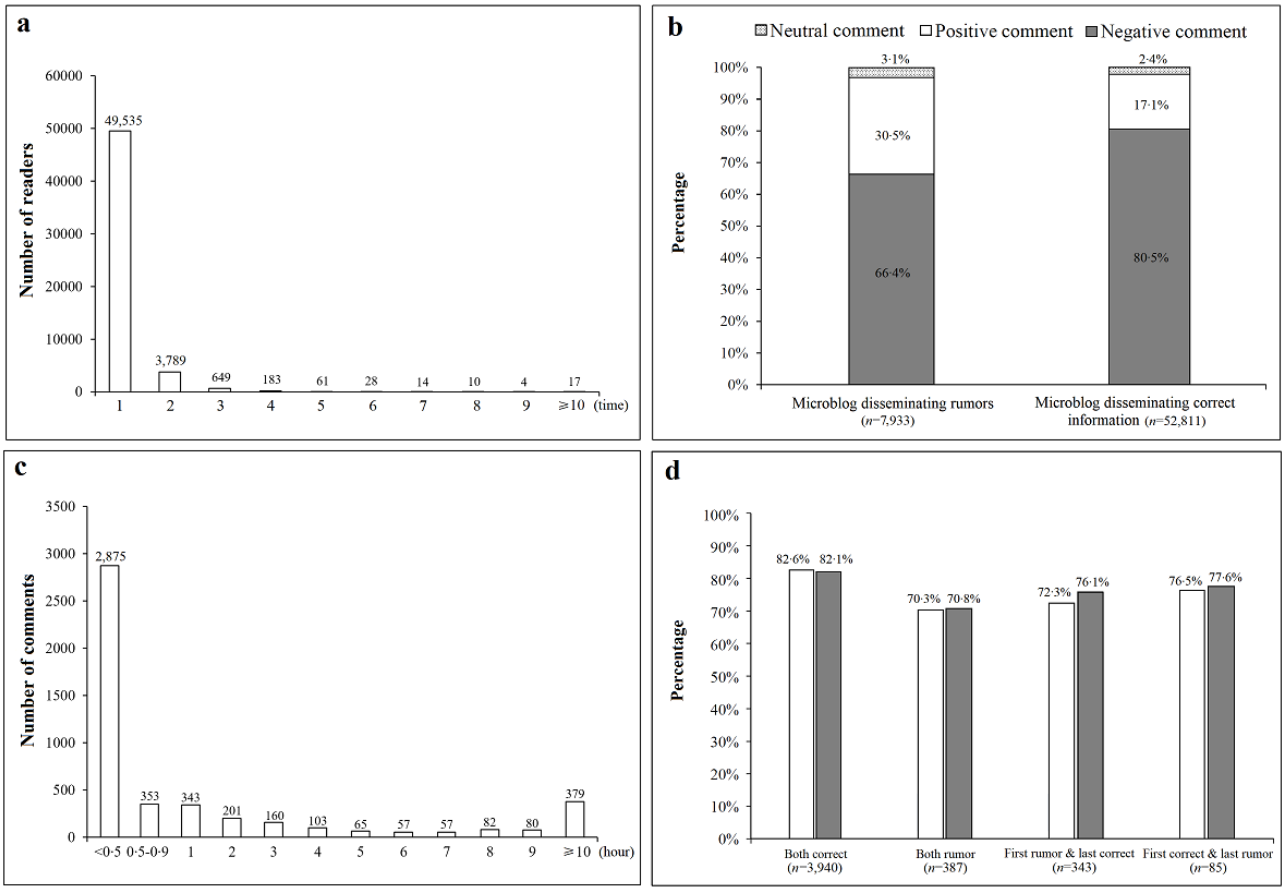
The proportion of rumor reports significantly decreased after the rumors were refuted by health authorities for case A (before: 534/666, 80.2%; after: 5685/45,657, 12.5%), case C (before: 1978/2101, 94.1%; after: 908/3796, 23.9%), and case D (before: 120/860, 14.0%; after: 192/2431, 7.9%) (P<.001) (Table 1); however, the proportion of rumor reports increased from 10.8% (50/463) to 60.5% (12,004/19,846) for case B after the official clarification with correction information was released (P<.001).

For case A (the novel coronavirus can be prevented with "Shuanghuanglian"), we identified a total of 60,744 comments posted by 54,290 microblog readers in response to 42 articles disseminating rumors and 78 articles refuting rumors and providing accurate information. Most readers posted once (49,535/54,290, 91.2%) or twice (3789/54,290, 7.0%) (Figure 3a), and most readers' comments expressed emotions with a negative connotation (eg, anger, anxiety, fear) in response to both types of microblog articles. However, the proportion of comments with negative emotions for articles disseminating correct information was significantly higher than the proportion for articles disseminating rumors (correct: 42,485/52,811, 80.5%; rumor: 5271/7933, 66.4%, χ_1^2=804.55, P<.001) (Figure 3b). Among readers who posted multiple comments, 60.5% (2875/4755) posted their first and last comments within half an hour (Figure 3c), and the proportion of negative comments posted by those readers did not change significantly when comparing their first versus last posts, no matter what type of article they read (both correct: χ_1^2=0.315, P=.58; both rumors: χ_1^2=0.025, P=.88; first rumor and last correct: χ_1^2=1.287, P=.26; first correct and last rumor: χ_1^2=0.033, P=.86) (Figure 3d).

Table 1. Media reports related to 4 most influential rumor cases before and after refuting the rumor.

Case and time period	Rumor report, n (%)	Correct report, n (%)	*P* value
Case A: The novel coronavirus can be prevented with "Shuanghuanglian"			<.001
Prerefutation	534 (80.2)	81 (12.2)	
Refuting the rumor (January 31, 11 PM to midnight)	4445 (59.5)	2079 (27.8)	
Postrefutation	5685 (12.5)	36006 (78.9)	
Case B: The novel coronavirus is the SARS coronavirus			<.001
Prerefutation	50 (10.8)	412 (89.0)	
Refuting the rumor (February 9, 9 PM to 10 PM)	3848 (93.2)	86 (2.1)	
Postrefutation	12004 (60.5)	5444 (27.4)	
Case C: Medical supplies intended for assisting Hubei were detained by the local government			<.001
Prerefutation	1978 (94.1)	3 (0.1)	
Ongoing (February 10, 11 PM to midnight)	1057 (87.8)	51 (4.2)	
Postrefutation	908 (23.9)	2721 (71.7)	
Case D: Asymptomatically infected persons were regarded as diagnosed COVID-19 patients with symptoms in official counts			<.001
Prerefutation	120 (14.0)	660 (76.7)	
Refuting the rumor (March 22, 7 AM to 8 AM)	11 (7.6)	126 (86.9)	
Postrefutation	192 (7.9)	2171 (89.3)	

Figure 3. Readers' comments in response to microblog articles for the case A rumor, "the novel coronavirus can be prevented with 'Shuanghuanglian'": (a) frequency of comments, (b) comment distribution by type of article, (c) time interval between the first and last comment for the same readers, and (d) proportion of negative comments in the first and last comments by type of article.



Discussion

Principal Findings

This study yielded 3 key findings. First, the contents and the number of rumors were strongly associated with the development of the COVID-19 epidemic in China, both temporally and geographically, and reports from personal social media accounts accounted for 71.7% (7648/10,664), 92.4% (14,696/15,902), 99.2% (3911/3943), and 99.7% (322/323) for cases A, B, C, and D, respectively, of all rumor reports. Second, releasing news to correct rumors decreased the dissemination of media articles for 3 of the 4 selected rumors (cases A, C, and D) but did not reduce news disseminations for case B. Third, the likelihood that readers posted negative comments in response to rumor reports and to reports disseminating correct information differed insignificantly (both correct reports: χ_1^2=0.315, *P*=.58; both rumors: χ_1^2=0.025, *P*=.88; first rumor and last correct report: χ_1^2=1.287, *P*=.26; first correct report and last rumor: χ_1^2=0.033, *P*=.86).

Interpretation of Findings

The findings of previous studies, regarding fluctuations in the number of rumors reported on a daily basis, the geographic variations for those reports across China [14], and major topics of the COVID-19 rumors [14,15], were generally concordant with our findings. The strong and positive correlation between reported rumors and the development of the COVID-19 epidemic likely reflects extreme tension, fear, and anxiety about the novel contagious disease among the population, especially residents of the regions most affected by the epidemic [16], such as Hubei, Zhejiang, and Guangxi. The likely reason for this was that the rapidly growing epidemic provoked emotional responses among inhabitants, leading them to scramble for any information they could find to alleviate their anxieties and fears. When the residents in those regions were unable to obtain timely official answers about major public concerns related to the COVID-19 epidemic, rumors emerged and were disseminated quickly [17]. For example, during the early stages of the COVID-19 epidemic in China, the question of how to dispose of used face masks emerged as a major public health concern. The question aroused 2031 microblog posts, but only 10 official Sina Microblog accounts released authoritative information on the issue before January 23, 2020 [18]. As a result, a total of 14 versions of misleading rumors emerged immediately on this topic, and 4740 articles disseminated improper strategies to dispose of discarded masks, such as heating them in the oven, scrubbing them with water, or cutting them into pieces [18].

This study presents a unique finding that personal social media accounts were the dominant source of rumor reports. This may reflect the combined effect of flourishing personal social media accounts [19], an inability for average citizens to detect misinformation [20], and amplification effects through (ie, an "echo chamber"—people are susceptible to peer influences when engaging in a homogeneous online cluster with others who share similar interests) [21].

Unexpectedly, we found that websites or news-based social media accounts were the second most common source of rumor reports. This may be a result of inadequate fact-checking of contents by the news industry or deliberate reports designed to attract public attention to these news media outlets [22,23]. As an example, the most widely spread rumor (Case A, the novel coronavirus can be prevented with "Shuanghuanglian," a traditional Chinese medicine used to treat the common cold) was first released by Weibo of Xinhua, an official news platform with nearly 100 million subscribers [24]. That release quickly led to a Sina Microblog topic titled "The novel coronavirus can be prevented with Shuanghuanglian," attracting 1.55 billion readings and 423,000 reader comments [25] and leading to panic buying of Shuanghuanglian.

The release of news to correct rumors was successful in reducing the spread of rumors in cases A, C, and D. This confirms that it is possible to mitigate the spread of rumors through authoritative releases of correct information in a timely manner [26]. Case B did not follow the same pattern, probably due to an unintentional slip of tongue by a government spokesperson at the 19th press conference in Hubei province on February 9, 2020 [27]. The spokesperson unintentionally and incorrectly confirmed the rumor, leading to a rapid increase in the number of rumor reports concerning Case B (from 41 to 15,767; Figure S5 in Multimedia Appendix 1) within 3 hours of the start of the press conference.

Strikingly, most reader comments in response to articles were negative, whether the articles reported rumors or correct information, and the proportions of negative comments were similar between both types of articles. These results match those of previous reports [28,29] and may reflect the fact that official correction news tends to disseminate correct information but largely overlooks any attempt to provide emotional support or empathy to affected people. Such factual presentation without empathy leads to unproductive dialogue with the emotional public, who release their negative sentiments in writing [29].

Implications

Our findings have 3 major implications. First, they underscore the importance of releasing scientifically accurate information in response to major public health concerns in a timely manner. Such releases help prevent the spread of rumors. To be effective, they require close cooperation between government officials, scientists, and the media.

In emerging health crises, such as the COVID-19 pandemic, public concerns inevitably arise that cannot always be answered immediately with accurate scientific evidence. In the absence of guidelines to satisfy public angst and curiosity for critical public health information while science progresses, basic guidance might be established by the WHO or academic experts to direct the release of available information through multiple channels, including government websites, news websites, and official and personal social media accounts. Such releases might reduce tension and anxiety among the public.

Second, the official release of correction articles tended to reduce rumor reports in China, suggesting similar effects could be expected in other countries. Success with this strategy would

require a real-time web-based monitoring system to detect rumor reports and close cooperation between government, media companies, and public health experts to release authoritative correction articles that refute rumors, with the caveat that such a strategy may not work well in countries where government and public health experts provide conflicting advice, such as in the United States for much of the situation with COVID-19 in 2020 [30].

Last, our findings suggest rigorous public health research is needed to generate solutions to unsolved questions, including how to effectively detect rumor reports as early as possible, which solutions might be best for releasing and disseminating correction articles, and how to effectively alleviate negative sentiments held by the public. In addition, the effectiveness of alternative postrumor approaches such as fact-checking and retractions (eg, deleting posts) should be evaluated.

Limitations

This study has limitations. First, our data collection approach likely underestimated the number of COVID-19 rumors, by missing some comments in response to Sina Microblog topics, since some were likely quickly detected and deleted by platform providers as part of efforts to control rumors [31]. Second, the lack of detailed data limited our ability to explore why most reader comments in response to media articles that disseminated correct information were negative and how to alleviate such negative sentiments to promote public support on fighting the COVID-19 pandemic. Further research is needed to address this critical public health matter. Third, we could not determine if any rumors we studied were published artificially by robots. As readers generally do not know who posted the reports, the role of robots is unlikely to influence our findings substantially. However, future studies may collect data on readers' responses to rumors spread by robots versus those spread by humans to quantify potential differences in their social impact and to develop specific interventions to deal with them separately.

Conclusion

The number of rumors about COVID-19 fluctuated greatly in China during the early months of the pandemic. The frequency of rumors was highly correlated with the magnitude of the epidemic. Rumor-disseminating reports emerged primarily from personal social media accounts, and releasing rumor-correcting news substantially reduced the spread of rumor reports. Reader comments in response to media articles disseminating rumors and to articles disseminating correct information expressed nearly the same percentage of negative emotions.

Individual governments and relevant international organizations such as the WHO should take immediate actions to develop regulations or guidelines to help social media companies and users of social media prepare and release reports properly and efficiently to prevent and control rumors, especially when scientific knowledge is limited during a health crisis such as the COVID-19 pandemic. In the meantime, our results suggest refutation reports were generally successful in reducing spread of rumors, and governments should encourage timely release of correct information to minimize the impact of rumors.

Acknowledgments

This work was funded by the Major Program of the National Social Science Foundation of China (20&ZD120). The sponsors had no role in the design and conduct of the study; analysis, report, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors' Contributions

GH conceptualized and designed the study, supervised the implementation of this study, and finalized the manuscript. JL and MZ collected and verified the underlying data. PC and JL carried out data analyses. PN and PC developed the manuscript. JL, PL, ML, and ZZ contributed to finding interpretation and revised the manuscript. DCS and YY critically edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 1153 KB - jmir_v23i12e27339_app1.docx]

References

1. Risk communication and community engagement (RCCE) readiness and response to the 2019 novel coronavirus (2019-nCoV): Interim guidance. World Health Organization. 2020 Jan 26. URL: https://apps.who.int/iris/bitstream/handle/10665/330678/9789240000773-eng.pdf?sequence=1&isAllowed=y [accessed 2020-11-08]
2. Walter N, Cohen J, Holbert RL, Morag Y. Fact-checking: a meta-analysis of what works and for whom. Polit Commun 2019 Oct 24;37(3):350-375. [doi: 10.1080/10584609.2019.1668894]
3. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]

4. Yang AS. Misinformation corrections of corporate news: corporate clarification announcements. Pac-Basin Financ J 2020 Jun;61:101315. [doi: 10.1016/j.pacfin.2020.101315]
5. Thorson E. Belief echoes: The persistent effects of corrected misinformation. Polit Commun 2015 Nov 19;33(3):460-480. [doi: 10.1080/10584609.2015.1102187]
6. Young DG, Jamieson KH, Poulsen S, Goldring A. Fact-checking effectiveness as a function of format and tone: evaluating FactCheck.org and FlackCheck.org. J Mass Commun Q 2017 Jul 06;95(1):49-75. [doi: 10.1177/1077699017710453]
7. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? an experimental evaluation of the effects of corrective information. Vaccine 2015 Jan;33(3):459-464. [doi: 10.1016/j.vaccine.2014.11.017] [Medline: 25499651]
8. Coronavirus disease (COVID-19) advice for the public: mythbusters. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2020-11-08]
9. Galvão J. COVID-19: the deadly threat of misinformation. Lancet Infect Dis 2021 May;21(5):e114 [FREE Full text] [doi: 10.1016/S1473-3099(20)30721-0] [Medline: 33031753]
10. Chinese Internet rumor clarification platform. Cyberspace Administration of China, Xinhua news. URL: http://www.piyao.org.cn/2020yqpy/ [accessed 2020-11-10]
11. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
12. Public health surveillance for COVID-19: interim guidance. World Health Organization. URL: https://apps.who.int/iris/handle/10665/337897 [accessed 2021-12-13]
13. Zhu B, Zheng X, Liu H, Li J, Wang P. Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos Solitons Fractals 2020 Nov;140:110123 [FREE Full text] [doi: 10.1016/j.chaos.2020.110123] [Medline: 32834635]
14. Islam MS, Sarkar T, Khan SH, Mostofa Kamal A, Hasan SMM, Kabir A, et al. COVID-19-related infodemic and its impact on public health: a global social media analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629 [FREE Full text] [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
15. Bastani P, Hakimzadeh SM, Bahrami MA. Designing a conceptual framework for misinformation on social media: a qualitative study on COVID-19. BMC Res Notes 2021 Nov 02;14(1):408 [FREE Full text] [doi: 10.1186/s13104-021-05822-2] [Medline: 34727969]
16. Liu M, Zhang H, Huang H. Media exposure to COVID-19 information, risk perception, social and geographical proximity, and self-rated anxiety in China. BMC Public Health 2020 Nov 04;20(1):1649 [FREE Full text] [doi: 10.1186/s12889-020-09761-8] [Medline: 33148201]
17. Larson H. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr 9;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
18. Voice express. ZhongQing HuaYun web-based public service platform. URL: http://www.cyyun.com/index.htm [accessed 2020-11-26]
19. The WeChat employment impact report 2019-2020. China Academy of Information and Communications Technology. URL: http://www.caict.ac.cn/kxyj/qwfb/ztbg/202005/t20200514_281774.htm [accessed 2020-11-26]
20. Scheufele DA, Krause NM. Science audiences, misinformation, and fake news. Proc Natl Acad Sci U S A 2019 Apr 16;116(16):7662-7669 [FREE Full text] [doi: 10.1073/pnas.1805871115] [Medline: 30642953]
21. Choi D, Chun S, Oh H, Han J, Kwon TT. Rumor propagation is amplified by echo chambers in social media. Sci Rep 2020 Jan 15;10(1):310 [FREE Full text] [doi: 10.1038/s41598-019-57272-3] [Medline: 31941980]
22. Oh O, Agrawal M, Rao H. Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. MIS Q 2013 Feb 2;37(2):407-426 [FREE Full text] [doi: 10.25300/misq/2013/37.2.05]
23. Sheldon P, Rauschnabel P, Honeycutt J. The Dark Side of Social Media: Psychological, Managerial, and Societal Perspectives. New York: Academic Press; 2019.
24. Novel coronavirus can be suppressed by novel shuanhuanglian oral liquid jointly discovered by Shanghai Pharmaceutical Institute and Wuhan Virus Institute. Xinhua viewpoint. URL: https://m.weibo.cn/1699432410/4466927503734764 [accessed 2020-11-26]
25. Novel coronavirus can be inhibited by shuanghuanglian. Weibo. URL: https://s.weibo.com/weibo?q=%23双黄连可抑制新型冠状病毒%23 [accessed 2020-11-26]
26. Huang H, Chen Y, Ma Y. Modeling the competitive diffusions of rumor and knowledge and the impacts on epidemic spreading. Appl Math Comput 2021 Jan 01;388:125536 [FREE Full text] [doi: 10.1016/j.amc.2020.125536] [Medline: 32834190]
27. What's the exactly the relationship between Novel Coronavirus and SARS. Beijing News. URL: http://www.bjnews.com.cn/news/2020/02/10/687056.html [accessed 2020-11-28]
28. 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]

29. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, et al. Debunking in a world of tribes. PLoS One 2017;12(7):e0181821 [FREE Full text] [doi: 10.1371/journal.pone.0181821] [Medline: 28742163]
30. Gonsalves G, Yamey G. Political interference in public health science during covid-19. BMJ 2020 Oct 06;371:m3878. [doi: 10.1136/bmj.m3878] [Medline: 33023874]
31. The coronavirus outbreak is a petri dish for conspiracy theories. Wired. URL: https://www.wired.com/story/coronavirus-conspiracy-theories/#main-content [accessed 2020-11-28]

Abbreviations

CYYUN: ZhongQing HuaYun web-based public service platform
WHO: World Health Organization

Edited by C Basch; submitted 21.01.21; peer-reviewed by J Glass, G Goh, H Moon; comments to author 06.04.21; revised version received 27.04.21; accepted 15.11.21; published 23.12.21.

Please cite as:
Ning P, Cheng P, Li J, Zheng M, Schwebel DC, Yang Y, Lu P, Mengdi L, Zhang Z, Hu G
COVID-19–Related Rumor Content, Transmission, and Clarification Strategies in China: Descriptive Study
J Med Internet Res 2021;23(12):e27339
URL: https://www.jmir.org/2021/12/e27339
doi:10.2196/27339
PMID:34806992

Original Paper

Examining Twitter Discourse on Electronic Cigarette and Tobacco Consumption During National Cancer Prevention Month in 2018: Topic Modeling and Geospatial Analysis

Jiahui Lu[1], PhD; Edmund W J Lee[2], PhD

[1]School of New Media and Communication, Tianjin University, Tianjin, China

[2]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Jiahui Lu, PhD
School of New Media and Communication
Tianjin University
No 92 Weijin Road
Tianjin, 300072
China
Phone: 86 18222418810
Email: lujiahui@tju.edu.cn

Abstract

Background: Examining public perception of tobacco products is critical for effective tobacco policy making and public education outreach. While the link between traditional tobacco products and lung cancer is well established, it is not known how the public perceives the association between electronic cigarettes (e-cigarettes) and lung cancer. In addition, it is unclear how members of the public interact with official messages during cancer campaigns on tobacco consumption and lung cancer.

Objective: In this study, we aimed to analyze e-cigarette and smoking tweets in the context of lung cancer during National Cancer Prevention Month in 2018 and examine how e-cigarette and traditional tobacco product discussions relate to implementation of tobacco control policies across different states in the United States.

Methods: We mined tweets that contained the term “lung cancer” on Twitter from February to March 2018. The data set contained 13,946 publicly available tweets that occurred during National Cancer Prevention Month (February 2018), and 10,153 tweets that occurred during March 2018. E-cigarette–related and smoking-related tweets were retrieved, using topic modeling and geospatial analysis.

Results: Debates on harmfulness (454/915, 49.7%), personal experiences (316/915, 34.5%), and e-cigarette risks (145/915, 15.8%) were the major themes of e-cigarette tweets related to lung cancer. Policy discussions (2251/3870, 58.1%), smoking risks (843/3870, 21.8%), and personal experiences (776/3870, 20.1%) were the major themes of smoking tweets related to lung cancer. Geospatial analysis showed that discussion on e-cigarette risks was positively correlated with the number of state-level smoke-free policies enacted for e-cigarettes. In particular, the number of indoor and on campus smoke-free policies was related to the number of tweets on e-cigarette risks (smoke-free indoor, $r49=0.33$, $P=.02$; smoke-free campus, $r49=0.32$, $P=.02$). The total number of e-cigarette policies was also positively related to the number of tweets on e-cigarette risks ($r49=0.32$, $P=.02$). In contrast, the number of smoking policies was not significantly associated with any of the smoking themes in the lung cancer discourse ($P>.13$).

Conclusions: Though people recognized the importance of traditional tobacco control policies in reducing lung cancer incidences, their views on e-cigarette risks were divided, and discussions on the importance of e-cigarette policy control were missing from public discourse. Findings suggest the need for health organizations to continuously engage the public in discussions on the potential health risks of e-cigarettes and raise awareness of the insidious lobbying efforts from the tobacco industry.

(J Med Internet Res 2021;23(12):e28042) doi:10.2196/28042

KEYWORDS

electronic cigarette; smoking; lung cancer; Twitter; national cancer prevention month; policy; topic modeling; cessation; e-cigarette; cancer; social media; eHealth; cancer prevention; tweets; public health

Introduction

Background

Tobacco control has been identified as a global public health priority by the World Health Organization [1]. Tobacco use is one of the leading causes of preventable deaths globally, and it is responsible for 7 million deaths worldwide [2]. In the United States, the Centers for Disease Control and Prevention (CDC) reported that tobacco use accounted for 1 in 5 deaths, leading to more than 480,000 deaths annually [3]. As such, for tobacco control, understanding public perception of tobacco products and their severe health risks, such as lung cancer, is essential to inform educational campaigns and tobacco control policies.

Particularly, health education campaigns and control policies should pay attention to tobacco consumption trends. In terms of traditional tobacco use, the CDC reported that smoking among adults had declined from 20.9% in 2005 to 13.7% in 2018, and the proportion of smokers who reported quitting had increased [4]. But, the popularity of emerging tobacco products such as electronic cigarettes (e-cigarettes) has been rising steadily since 2010. E-cigarettes are electronic devices that are used to deliver nicotine and other chemicals to users through inhalable aerosols. The US Surgeon General declared that e-cigarette use was an epidemic among youth in 2018, given that 21% of high school seniors in the United States reported using e-cigarettes in the preceding 30 days in the same year [5]. In particular, Juul captured and dominated 73% of the e-cigarette market through its product promotion featuring youth culture and lifestyle on different social media platforms [6].

The overwhelming popularity of e-cigarettes may be due to conflicting messages in the public. Some argue that e-cigarettes could help with smoking cessation as they appear to pose fewer health risks than traditional cigarettes [7-9]. Meanwhile, others cautioned against e-cigarette consumption as they contain nicotine, a highly addictive drug, that can lead to the use of other tobacco products [10-12]. Even though there is no established connection between e-cigarette use and severe diseases such as lung cancer, there is evidence of lung injuries associated with its use [13-15]. Notably, the CDC declared an outbreak of lung injuries associated with e-cigarette use in 2019 [16].

Related Work

Survey studies have investigated the perceived associations of smoking traditional tobacco products and e-cigarettes with health diseases such as lung cancer. Smoking is recognized as a major risk factor for lung cancer by the public [17]. In contrast, the public may not view e-cigarettes as a likely cause of lung cancer [18,19] but, instead, may view e-cigarettes as a safe alternative to combustible cigarettes [20,21]. There are very few studies that have compared public perceptions of both smoking and vaping in relation to severe health consequences using social media data. Currently, a growing body of social media studies [22,23] on e-cigarettes and smoking show that e-cigarettes are often presented as a positive and healthier alternative to smoking, despite controversies surrounding their effectiveness in smoking cessation and the likelihood of initiating adolescents to consuming other tobacco products. Nevertheless, there is limited evidence about the extent that the public perceives lung cancer to be a health consequence of both smoking and vaping [24-26]. Such comparisons are critical as they can highlight contextual, psychological, and behavioral factors specific to the reasons for which individuals consume different tobacco products. This knowledge is valuable as it can inform strategic messaging and educational programs to facilitate behavioral changes in reducing tobacco consumption [27].

Furthermore, to the best of our knowledge, none of the existing social media studies has investigated spatial patterns of tobacco conversations on social media in relation to implementation of health policies. For example, it has been found that the number of obesity-related policies in certain geographic regions were associated with an increase in obesity prevention discussions on Twitter within the same area [28]. This likely suggests that the number of health policies, or the policy environment in general, have a reciprocal relationship with consumers' health awareness. Putting this in the context of tobacco control, the number of state-level tobacco policies in the United States may be a reflection of the risk perception of tobacco products and e-cigarettes, and this heightened awareness of risk on a societal level may be impetus for policy makers to enact more laws to rein in tobacco consumption. Therefore, it is crucial to study how tobacco policies of the various states are associated with public perceptions and discussions of health effects of tobacco products on social media.

Tobacco Discourse During US National Cancer Prevention Month

While the majority of social media studies largely examine public discourse of e-cigarette and smoking in general, this study examined tobacco discourse in the context of US National Cancer Prevention Month, which is an annual campaign led by the American Institute for Cancer Research in the month of February that aims to foster cancer knowledge and promote cancer prevention practices [29]. Particularly, lung cancer, which is a deadly disease that is caused by tobacco consumption [30], is one of prominent cancer topics during the campaign. Past research has shown that month-long cancer campaigns, such as National Cancer Prevention Month, are the key to increasing awareness of cancer and its associated risk factors [31]. As such, National Cancer Prevention Month offer opportunities to examine e-cigarette and smoking discourse related to lung cancer.

Past research has shown that public discourse during cancer campaigns may be different from that during other months. Cancer campaigns raise public awareness by promoting cancer conversations about risk factors and preventions on Twitter [32]. Also, public discourse during cancer campaigns is often driven by health organizations, advocacy groups, and influential personalities such as celebrities [32,33]. The agenda of cancer campaigns typically involves disseminating cancer education messages, advocating prevention engagement, and sharing affective stories of survivors [33,34]. In addition to these messages, Twitter users could respond and selectively follow and express their own opinions on lung cancer. As such, public discourse during campaigns would likely reflect how the public

interact with official messages on tobacco consumption and lung cancer. This knowledge can be important for public health officials in identifying gaps in health education.

Objective

First, we aimed to examine if National Cancer Prevention Month plays a role in promoting conversations on the link of e-cigarettes and smoking with lung cancer. Second, we aimed to examine and compare public discourse in the United States on smoking and e-cigarette in the context of lung cancer on Twitter during National Cancer Prevention Month. Third, we aimed to examine if there were spatial patterns of smoking and e-cigarette's themes in the United States during National Cancer Prevention Month. Fourth, we explored the relationship between e-cigarette and smoking discussions on Twitter during National Cancer Prevention Month with implementation of tobacco control policies. As such, we put forth 4 research questions: (1) Does national cancer prevention month promote e-cigarette and smoking conversation related to lung cancer? (2) What are the key themes in e-cigarette and smoking tweets within the broader context of lung cancer discussion during National Cancer Prevention Month? (3) Are there geospatial differences in how e-cigarette and smoking tweets were distributed across the United States during National Cancer Prevention Month? (4) What is the relationship between the number of tobacco control policies in states and themes of e-cigarette and smoking tweets?

Methods

Data Collection

Data were retrieved from an existing data set of US English-based lung cancer tweets that contained the term "lung cancer" purchased from Twitter. A list of 28 keywords, such as "e-cigarette," "vape," and "juul" [35,36], and another list of 21 keywords, such as "cigarette" and "smoking" [37], were developed and used to extract e-cigarette and smoking tweets from the data set, respectively (Table 1). Tweets that contained both sets of keywords were excluded from further analysis to obtain a proper comparison.

Table 1. Search keywords for data collection.

Topic	Keywords
E-cigarette	electronic cigarette; vap*; e-cig*; ecig*; e cig; e-pen; epen; e pen; e-juice; ejuice; e juice; e-liquid; eliquid; e liquid; esmoke; e-smoke; e smoke; e-hookah; ehookah; e hookah; e-pipe; epipe; e pipe; atomizer; juul; njoy; v2 cig; joye510
Smoking	cig*; tobacco; waterpipe; water pipe; hooka; smok*; chew; nicotine; shisha; sheesha; bidi; beedi; kretek; narghile; argileh; cheroot; snuff; snus; betel; gutkha; toombak

Data Processing and Analysis

R statistical software (The R Project) was used for textual analyses. Data were preprocessed and cleaned before advanced textual analysis. Texts were formatted to lower case. Different forms of phrases that had the same meaning were transformed into a common format to facilitate future text processing, such that "e cig," "e-cig," "ecig," and "electronic cigarette" were reformatted into "ecigarette." Common English stop words, such as "the" and "of," special characters, and punctuations were removed. The remaining texts were tokenized and lemmatized to further avoid inflected words. The word "lung cancer" was also removed from the analysis.

Topic Modeling

Topic modeling, using latent Dirichlet allocation, was employed to understand the differences between themes of e-cigarette and smoking in the context of lung cancer discourse. Latent Dirichlet allocation is a popular and widely used algorithm for topic modeling, by which documents are modeled as mixtures over topics and a topic is characterized as a distribution of words [38]. R software (topicmodels package [39]) was used.

The latent Dirichlet allocation algorithm requires a predefined number, *k*, of topics, which we determined with the perplexity and log-likelihood indices. Both indices have been conventionally used to evaluate the model [38,40]. A lower perplexity score and a higher log-likelihood score indicate better generalization performance. We trained topic models from k=2 to k=20. To maximize the diversity of discussion while at the same time minimize topic overlaps, we selected the k model when the $k+1$ model did not improve [27,38]. After the topics were generated, we named each topic by examining keywords and posts that were representative of those topics. Topics deemed to have similar themes were combined to facilitate discussions.

To evaluate the prevalence of topics, we used methods described in [27] to determine the topic of each tweet. The output of topic modeling included estimates of the proportion of each topic present in each tweet. Based on this output, we assigned each tweet the topic that had the highest predicted proportion. For example, the tweet "February 7, 2018. The day everyone who juuls simultaneously got lung cancer" was estimated to be 1.1% similar to other tweets under the topic *affective reasoning*, 1.1% similar to other tweets under the topic *personal experiences*, 95.6% similar to other tweets under the topic *sarcasm*, 1.1% similar to other tweets under the topic *cognitive reasoning*, and 1.1% similar to other tweets under the topic *e-cigarette risk* and was, therefore, classified under the topic *sarcasm*.

We compared the temporal distribution of e-cigarette and smoking tweets using chi-square analysis.

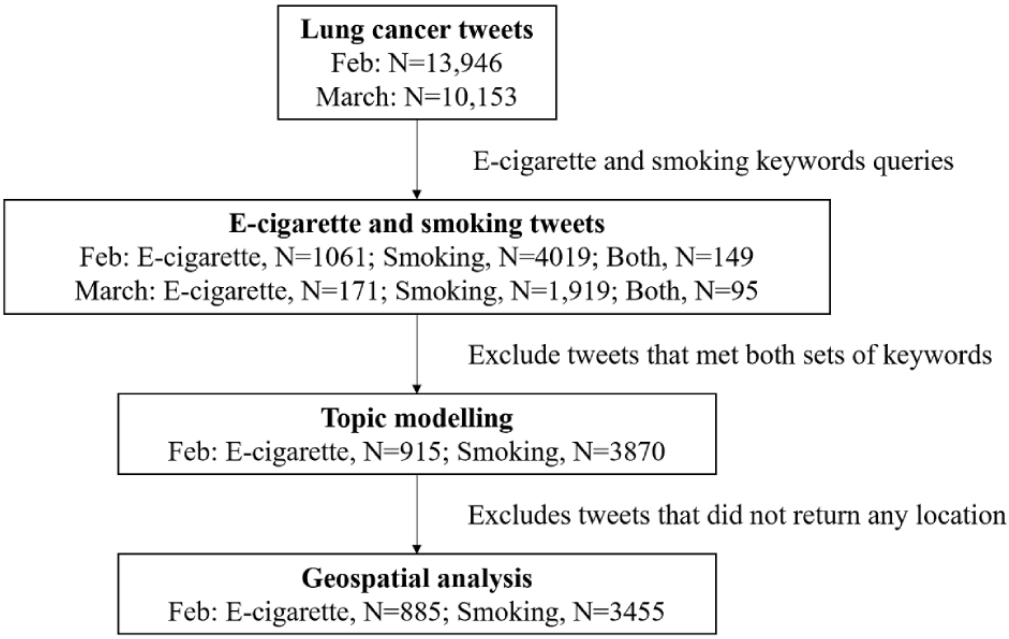
Geospatial Analysis

The geolocation of a tweet was determined by the self-reported location in the profile of the relevant Twitter user. We imported the location strings into the Google Maps geocoordinates application programming interface (API) to obtain the geocoordinates and the corresponding states in the United States. Then, we manually checked to ensure the state information is correct for each tweet. Tweets that did not have a user-reported location or whose reported location string did not return any results were excluded from the geospatial analysis. To further

understand the spatial distribution of themes, ratio values were calculated for each state by dividing the number of tweets in each theme by the total number of tweets in each state. We plotted the ratio values by state to visualize the spatial distribution patterns.

To further understand how state policies might affect twitter discussions of e-cigarette and smoking, we compared the number of tweets with the number of state policies (existing, introduced, or recently enacted in the first quarter of 2018) related to e-cigarettes and smoking. State policies were obtained from the tobacco use data portal from the US CDC [41]. For e-cigarettes, the data included policies related to the sale of e-cigarettes to youth, retail licenses to sell e-cigarette, smoke-free indoor, smoke-free on campus, taxes on e-cigarette products, and preemption. For smoking, the data included all policies similar to those for e-cigarettes and additional policies on fire safety. We conducted bivariate correlation analysis (Pearson correlation) of the total number of tweets and the total number of state policies, as well as between the themes of tweets and types of tobacco control policies.

Figure 1. Study flowchart.



Results

Temporal Distribution of E-cigarette and Smoking Tweets

The data set had 13,946 publicly available tweets obtained during National Cancer Prevention Month (ie, February) and 10,153 tweets obtained during March in 2018 in the United States. The keyword queries returned 1061 e-cigarette tweets and 4019 smoking tweets during National Cancer Prevention Month, and 171 e-cigarette tweets and 1919 smoking tweets during March (Figure 1). Of these, 149 tweets during National Cancer Prevention Month and 95 tweets during March contained both sets of keywords and were removed. This yielded 6.56% (915/13,946) e-cigarette tweets and 27.7% (3870/13,946) general smoking tweets made by 839 and 3501 unique users during National Cancer Prevention Month, and 0.75% (76/10,153) e-cigarette tweets and 18.0% (1824/10,153) general smoking tweets made by 57 and 1643 unique users during March.

There was a significant difference between the temporal distributions for e-cigarette and smoking tweets (χ^2=256.85, P<.001). This suggests that National Cancer Prevention Month did promote lung cancer discussion associated with e-cigarette and smoking.

E-cigarette and Smoking Tweets Themes

During National Cancer Prevention Month, 3 major e-cigarette themes—comprising 5 topics—emerged (Table 2). The topics *affective reasoning*, *cognitive reasoning*, and *sarcasm* were categorized under the overarching theme *e-cigarette debate*. First, Twitter users made the link between e-cigarettes and lung cancer based on emotional evaluation of their experiences or anecdotal stories they have heard. Second, there were those who discussed and cognitively processed e-cigarette information presented to them. For instance, one of the tweets highlighted that lung cancer could take years to develop and simply "juuling" for a short period may not be enough to develop lung cancer. Third, some of the Twitter users expressed sarcasm when discussing e-cigarettes, pushing back on the idea of "juuling" and lung cancer. These 3 topics accounted for half of all tweets (454/915, 49.7%). The second major theme was *e-cigarette risks* (145/915, 15.8%), where many tweets discussed scientific evidence on the risks of e-cigarette on lung diseases. The third major theme was *personal experience*s, which constituted approximately one-third of the tweets (316/915, 34.5%). Many tweets in this category encouraged others to stop vaping, with users citing stories they had heard about someone who contracted lung cancer by consuming e-cigarettes.

Table 2. E-cigarette themes from tweets (based on latent Dirichlet allocation algorithm).

Theme	Words	Examples	Tweets (n= 915), n (%)
E-cigarette debates			454 (49.7)
Affective reasoning	Juul, hit, kid, people, rumor, cause, shit, really, report, untrue	Drop your juuls like deadass. Have heard of three people my age who have been diagnosed with lung cancer from juuls. Feel like we all saw this coming	158 (17.3)
Cognitive reasoning	Juul, year, cause, cancer, develop, know, take, use, lung, say	When people think that juuls give you lung cancer but lung cancer takes years to develop	127 (13.9)
Sarcasm	Juul, everyone, day, February, simultaneously, friend, today, people, college, think	February 7, 2018. The day everyone who juuls simultaneously got lung cancer.	169 (18.5)
Personal experiences	Friend, cousin, stop, girl, good, sister, neighbor, immediately, son, sorority	STOP JUULING IMMEDIATELY, My best friends neighbors girl friend's sorority sister's cousin's step son got lung cancer from a single hit of juul. Drop these cancer sticks.	316 (34.5)
E-cigarette risks	Juul, link, lung, kid, disease, ecigarette, severe, flavoring, hit, addict	ecigarette Flavorings linked to Severe LUNG disease https://t.co/2VhStSkI0s #lungdisease #lungcancer #ecigarettes #cancer #ecigaretteflavoring #severlungdisease #howbadareecigarettes	145 (15.8)

Unlike themes expressed in e-cigarette tweets, which showed that users were divided over the association of e-cigarettes with lung cancer, those expressed in smoking tweets (Table 3) showed that users were largely unanimous in perceiving the link between smoking and lung cancer. More than half of tweets (2251/3870, 58.1%) were classified under the theme *policy discussion*, which equated the importance of tobacco control policies with that of gun control policies. Furthermore, more than 20% of tweets (843/3870) focused on the theme *smoking risks*—tweets promoted smoking cessation by mentioning scientific facts of smoking and its relation to lung cancer. Another major theme was *personal experiences*, with 20% of the tweets (776/3870) having stories of how users or their families suffered lung cancer because of smoking.

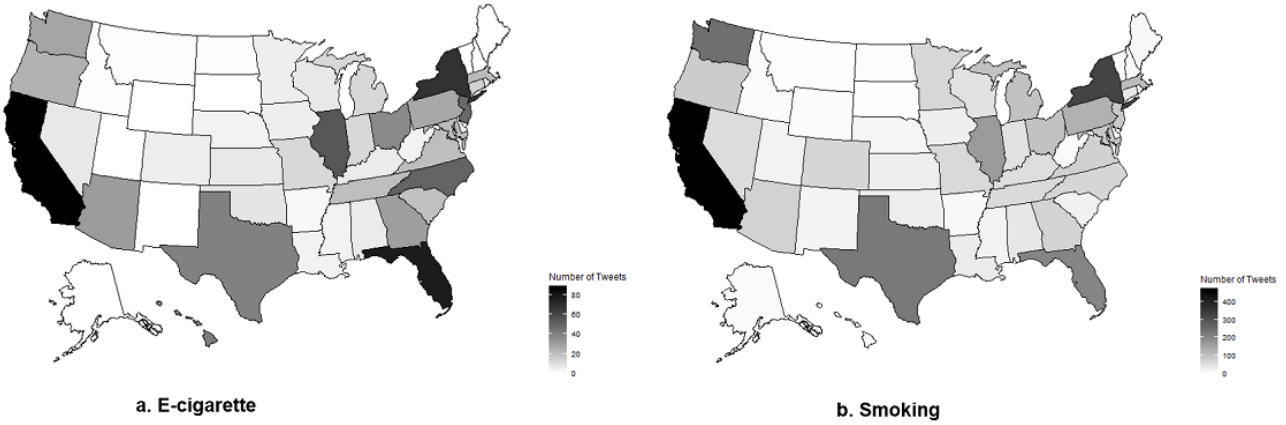
Table 3. Smoking themes (based on latent Dirichlet allocation) from tweets.

Theme	Words	Examples	Tweets (n= 3870), n (%)
Policy discussion			2251 (58.1)
Tobacco lobbying	Tobacco, kid, one, lose, prevent, explain, love, lobbyist, invite, this	This is like inviting tobacco lobbyists to explain to kids who have lost loved ones to lung cancer how we can prevent smoking deaths. Lobbyists don't deserve a seat at this table.	1831 (47.3)
Smoking control policy	Smoker, cigarette, gun, smoke, cause, blame, tobacco, gum, chew, death	Higher prices / taxes on cigarettes=less deaths due to lung cancer. Seat belts / stricter regulations on vehicle safety=less auto deaths	420 (10.8)
Smoking risks	Smoke, cancer, smoker, lung, non, die, people, risk, quit, cigarette	433 Americans die daily from #lungcancer. The majority of people living with lung cancer r nonsmokers or have quit smoking. Anyone, smoker or nonsmoker, can get lung cancer. While smoking greatly increases the risk of #lungcancer , NO ONE DESERVES CANCER. @theNCI #LCSM https://t.co/9CtilnJLzm	843 (21.8)
Personal experiences	Smoke, get, cigarette, people, die, tobacco, cause, someone, kill, make	Great! My big brother was smoking for the last fifty years Lung cancer finally killed him. Small pain under the arm one sunny morning. Two years later a 3cm tumor killed him.	776 (20.1)

Spatial Distributions of E-cigarette and Smoking Tweets and Themes

The geospatial analysis included 96.7% (885/915) of the e-cigarette tweets and 89.3% (3455/3870) of the smoking tweets during National Cancer Prevention Month. Overall, discussions of e-cigarettes and smoking in relation to lung cancer occurred mostly in the coastal areas and the eastern part of the country (Figure 2). California, New York, Florida, Texas, and Illinois had the most tweets about e-cigarettes and smoking. We focused our analysis on states that had more than 20 tweets in order to draw meaningful distribution patterns (Figures S1 and S2 in Multimedia Appendix 1).

Figure 2. The spatial distribution of e-cigarette and smoking tweets mentioning lung cancer during US national cancer prevention month (February 2018).



a. E-cigarette

b. Smoking

For e-cigarette tweets, California, Arizona, Pennsylvania, Illinois, and Virginia had more tweets debating e-cigarettes than tweets containing the other two themes. Oregon, Texas, Tennessee, and North Carolina had more tweets asking people to stop vaping based on personal stories. Hawaii and the Washington state had more tweets on scientific evidence showing the link between e-cigarettes and lung diseases.

For smoking tweets, most states predominantly had tweets about policy discussions regarding smoke control and how tobacco control policies were important to reign in tobacco companies, equivalent to how gun control policies would restrict gun lobbyists. Nevada and Kentucky had more tweets about the scientific evidence of smoking risks than those about the other two themes.

Association With State-Level Tobacco Control Policies

The number of tweets under the theme *e-cigarette risks* was positively associated with the number of both indoor arena and on campus smoke-free policies (smoke-free indoor: r_{49}=0.33, *P*=.02; smoke-free campus: r_{49}=.32, *P*=.02). Likewise, the number of tweets under the theme *e-cigarette risks* was positively associated with the total number of e-cigarette policies, r_{49}=.32, *P*=.02). There were no statistically significant associations between the 3 smoking themes and the number of smoking policies in the context of lung cancer (*P*>.13) (Tables S1 and S2 in Multimedia Appendix 2).

Discussion

Principal Findings

We examined the prevailing topics and distributions of discussions in Twitter about e-cigarettes and traditional tobacco consumption during the National Cancer Prevention Month in 2018 within the broader context of lung cancer to offer key insights on how the public perceives health risks of both e-cigarettes and smoking and potentially help public health organizations to be more strategic in their messaging and tobacco control efforts by targeting different tobacco products.

First, the findings of temporal distributions of e-cigarette and smoking tweets suggest that National Cancer Prevention Month promoted both e-cigarette and smoking conversations related to lung cancer. What we found notable was that National Cancer Prevention Month promoted e-cigarette conversations more than smoking conversations in the context of lung cancer. Without the cancer campaigns, lung cancer discourse on Twitter were rarely about e-cigarettes (76/10,153, 0.75%) and mostly revolved around the harms of smoking (1824/10,153, 18.0%). This is likely for a few reasons. In February 2018, the American Cancer Society [42] first released an official statement that discouraged youths or young adults from using any tobacco products including e-cigarettes. As such, it might have generated additional attention to the harms of e-cigarettes in relation to lung cancer during National Cancer Prevention Month. Second, 2018 National Youth Tobacco Survey data raised concerns about the vaping epidemic by showing an alarming surge in e-cigarette use among youths from 2017 to 2018; there was a 78% increase in e-cigarette use among high school students and a 48% increase among middle school students [43]. In addition, the nature of National Cancer Prevention Month itself, as a cancer awareness campaign, did indeed promote public debate and concerns about e-cigarettes on Twitter as demonstrated by our data.

Second, the findings of our thematic analysis suggest that Twitter users were aware of the risks of lung cancer from smoking but were split over the potential health effects of vaping. While some of the Twitter users evaluated the link between e-cigarettes and lung cancer based on personal experiences or anecdotal stories they have heard, others processed e-cigarette information in a more cerebral manner and were convinced of the health risks of e-cigarettes. This split in attitude toward e-cigarettes may be the result of mixed communication messaging from public health organizations. For instance, while the CDC has acknowledged the risks of e-cigarettes, particularly for young people due to the presence of nicotine, the long-term health effects of e-cigarettes have been debated [44]. And the National Academies of Sciences, Engineering, and Medicine [10] released a report concluding that e-cigarette use could not be strictly categorized as harmful or beneficial because it would require more long-term studies on the health effects of vaping.

Another significant finding was that themes of political lobbying and policy making were absent from e-cigarette tweets, but not from those about traditional tobacco consumption, during the

cancer campaign. When discussing smoking, Twitter users were mindful of the political lobbying by tobacco industries (and equated it to that of gun lobbyists), but this particular theme was missing from e-cigarette tweets. This is crucial, as it suggests that the political lobbying efforts by e-cigarette companies may not be as visible or prominent as those of the traditional tobacco industry. This is a cause for concern. After all, the tobacco industry is very much involved in the e-cigarette industry, as shown by the acquisition of Juul by Altria (formerly known as Philip Morris) for US $12.8 billion in 2018 [45]. Moreover, in recent years, e-cigarette companies have increased efforts in boosting scientific legitimacy in the context of health effects from consumption of their vaping products. For instance, Juul established the JLI Science lab, to fund scientific research on the effects of vaping products—a move that resembled the tobacco industries' use of research for political lobbying efforts in the 1980s [46]. In terms of policy actions, public debate on the need for tighter regulations over e-cigarettes is critical. Research has shown that supply-side restrictions—such as limiting tobacco retail outlet density—are effective in reducing tobacco consumption [47]. If there are any indications that the current ban on flavored e-cigarettes by the United States and other countries has an effect on the tobacco industry, Altria will revise terms of investments in Juul [48].

Third, this study demonstrated geospatial differences in e-cigarette and smoking discussions on Twitter during National Cancer Prevention Month. In terms of discussing e-cigarette risks, results showed that only 2 states—Hawaii and Washington—had more discussions than those of the others. In the state of Washington, 30% of 12th grade students used e-cigarettes [49], compared to the use of other tobacco products such as smokeless tobacco (4%) or cigars (7%). In the state of Hawaii, high school teenagers vaped twice as much as the national average [50]. To curb the vaping epidemic, Hawaii was one of the first states to raise the legal age of sales for tobacco from 18 years to 21 years, in an effort to prevent young people from nicotine addiction and the harms of tobacco use [51].

There was a positive correlation between discussion of e-cigarette risks and the number of smoke-free policies at the state level. While the findings cannot be used to make any causal claims, it is worth nothing that there may be a reciprocal relationship between public awareness of e-cigarette risks and the passing of smoke-free policies. In other words, when the public becomes aware of the risks of e-cigarettes, they may encourage local representatives to push for more smoke-free policies. At the same time, the passing of smoke-free policies may further increase awareness of e-cigarette risks in the general public.

Though discussion of tobacco risks and the number of smoke-free policies were not correlated, as discussed, people still mentioned the importance of smoking control policies in their tweets. The data suggest that policy engagement and public awareness and discussion of tobacco risks are symbiotic. When the risk awareness of a tobacco product is low, especially for emerging tobacco products such as e-cigarettes, public policy engagement motivated by the community leaders or public health organizations may heighten risk awareness. Once the public are adequately educated on the health risks of a tobacco product (eg, combustible cigarettes), this risk awareness may, in turn, fuel discussions on the need for stringent tobacco control policies, as well as strategically address tactics of political lobbying and messaging by the tobacco industry. In other words, the findings of our study suggest that public health organizations should focus on both improving risk awareness of tobacco products, as well as engaging and educating the public on the importance of tobacco control policies, because these strategies complement and reinforce one another.

We believe that our findings will be useful to help health communication scholars understand public perception and attitudes toward e-cigarettes and smoking. Future studies should (1) test potential reciprocal relationship between policy engagement and risk awareness of tobacco products; (2) investigate the underlying mechanisms, specifically examine how National Cancer Prevention Month or other cancer awareness months could promote e-cigarette discussions with randomized controlled trials, and identify the best strategies in educating the public about the harms of vaping; and (3) replicate our study by examining how cancer awareness months drive conversations about other cancers (eg, breast cancer, prostate cancer) compared with other noncancer awareness months and how various health policies (eg, health insurance) across different states are associated with cancer discussion.

Limitations

First, while social media sources, such as Twitter, can be used to gauge public opinion and sentiments toward smoking and e-cigarette, we are mindful that they may not be representative, and as such, there are constraints on the generalizability of the results. For example, our data came from publicly available posts, and thus, we were not able to capture themes and sentiments toward smoking and e-cigarettes in private posts. Also, because not all users reported their locations in their profiles, there may be potential selection biases in the geospatial analysis. In addition, we are cognizant that there potentially could be a spill-over effect because our data were collected from consecutive months in February and March. However, we are confident that this was not a major issue of concern given that the number of e-cigarette tweets in February (n=1061) during National Cancer Prevention Month was much greater than the number of e-cigarette tweets in March (n=171). Finally, we excluded tweets with both sets of keywords that might introduce bias and ran the same analyses; we found that the results did not substantially differ; therefore, we are confident that our results are robust.

Conclusion

The public is aware of smoking and lung cancer risks, but people were generally divided over the risks of e-cigarettes in relation to lung cancer. Public health organizations should invest in strategic messaging efforts over social media to address any misinformation about e-cigarettes because there is a reciprocal relationship between public awareness and discussion on tobacco products on social media and the implementation of tobacco control policies.

Acknowledgments

This project was supported by a research grant from Nanyang Technological University (grant M020060110).

Authors' Contributions

JL and EL contributed to the design of the study. JL performed the analyses. JL and EL drafted the manuscript. Both authors contributed to revision of the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The spatial distribution of Twitter themes in relation to lung cancer during US national cancer prevention month (February 2018).
[DOCX File , 268 KB - jmir_v23i12e28042_app1.docx]

Multimedia Appendix 2
Correlation between the tweet number of themes and the number of state-level tobacco policies.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v23i12e28042_app2.xlsx]

References

1. Tobacco control. World Health Organization. URL: https://www.who.int/tobacco/control/en/ [accessed 2021-12-13]
2. WHO report on the global tobacco epidemic, 2017: monitoring tobacco use and prevention policies. World Health Organization. URL: https://www.who.int/publications/i/item/9789241512824 [accessed 2021-12-13]
3. Tobacco-related mortality. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm [accessed 2021-12-13]
4. Current smoking among adults in 2018 (States). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm#states [accessed 2021-12-13]
5. Surgeon general's advisory on e-cigarette use among youth. Centers for Disease Control and Prevention. 2018 Dec. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/surgeon-general-advisory/index.html [accessed 2021-12-13]
6. Czaplicki L, Kostygina G, Kim Y, Perks SN, Szczypka G, Emery SL, et al. Characterising JUUL-related posts on Instagram. Tob Control 2020 Nov;29(6):612-617. [doi: 10.1136/tobaccocontrol-2018-054824] [Medline: 31266903]
7. Polosa R, Caponnetto P, Morjaria JB, Papale G, Campagna D, Russo C. Effect of an electronic nicotine delivery device (e-Cigarette) on smoking reduction and cessation: a prospective 6-month pilot study. BMC Public Health 2011 Oct 11;11:786 [FREE Full text] [doi: 10.1186/1471-2458-11-786] [Medline: 21989407]
8. Beard E, West R, Michie S, Brown J. Association between electronic cigarette use and changes in quit attempts, success of quit attempts, use of smoking cessation pharmacotherapy, and use of stop smoking services in England: time series analysis of population trends. BMJ 2016 Sep 13;354:i4645 [FREE Full text] [doi: 10.1136/bmj.i4645] [Medline: 27624188]
9. Zhuang Y, Cummins SE, Sun JY, Zhu S. Long-term e-cigarette use and smoking cessation: a longitudinal study with US population. Tob Control 2016 Oct;25(Suppl 1):i90-i95 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053096] [Medline: 27697953]
10. Committee on the review of the health effects of electronic cigarettes of electronic nicotine delivery systems. In: Stratton K, Kwan LY, Eaton DL, editors. Public Health Consequences of E-cigarettes. Washington, DC: National Academies Press; 2018.
11. Bold KW, Kong G, Camenga DR, Simon P, Cavallo DA, Morean ME, et al. Trajectories of e-cigarette and conventional cigarette use among youth. Pediatrics 2018 Jan 04;141(1):e20171832 [FREE Full text] [doi: 10.1542/peds.2017-1832] [Medline: 29203523]
12. Barrington-Trimis JL, Berhane K, Unger JB, Cruz TB, Urman R, Chou CP, et al. The e-cigarette social environment, e-cigarette use, and susceptibility to cigarette smoking. J Adolesc Health 2016 Jul;59(1):75-80 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.03.019] [Medline: 27161417]
13. Kalininskiy A, Bach CT, Nacca NE, Ginsberg G, Marraffa J, Navarette KA, et al. E-cigarette, or vaping, product use associated lung injury (EVALI): case series and diagnostic approach. Lancet Respir Med 2019 Dec;7(12):1017-1026. [doi: 10.1016/S2213-2600(19)30415-1] [Medline: 31711871]
14. Lee H, Park S, Weng M, Wang H, Huang WC, Lepor H, et al. E-cigarette smoke damages DNA and reduces repair activity in mouse lung, heart, and bladder as well as in human lung and bladder cells. Proc Natl Acad Sci U S A 2018 Feb 13;115(7):E1560-E1569 [FREE Full text] [doi: 10.1073/pnas.1718185115] [Medline: 29378943]

15. Rowell TR, Tarran R. Will chronic e-cigarette use cause lung disease? Am J Physiol Lung Cell Mol Physiol 2015 Dec 15;309(12):L1398-L1409 [FREE Full text] [doi: 10.1152/ajplung.00272.2015] [Medline: 26408554]
16. Outbreak of lung injury associated with the use of e-cigarette, or vaping, products. Centers for Disease Control Prevention. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2021-12-13]
17. Greillier L, Cortot AB, Viguier J, Brignoli-Guibaudet L, Lhomel C, Eisinger F, et al. Perception of lung cancer risk: impact of smoking status and nicotine dependence. Curr Oncol Rep 2018 Mar 05;20(Suppl 1):18. [doi: 10.1007/s11912-017-0650-1] [Medline: 29508085]
18. Baeza-Loya S, Viswanath H, Carter A, Molfese DL, Velasquez KM, Baldwin PR, et al. Perceptions about e-cigarette safety may lead to e-smoking during pregnancy. Bull Menninger Clin 2014 Sep;78(3):243-252 [FREE Full text] [doi: 10.1521/bumc.2014.78.3.243] [Medline: 25247743]
19. Pepper JK, Emery SL, Ribisl KM, Rini CM, Brewer NT. How risky is it to use e-cigarettes? smokers' beliefs about their health risks from using novel and traditional tobacco products. J Behav Med 2015 Apr 28;38(2):318-326 [FREE Full text] [doi: 10.1007/s10865-014-9605-2] [Medline: 25348584]
20. Adkison SE, O'Connor RJ, Bansal-Travers M, Hyland A, Borland R, Yong H, et al. Electronic nicotine delivery systems: international tobacco control four-country survey. Am J Prev Med 2013 Mar;44(3):207-215 [FREE Full text] [doi: 10.1016/j.amepre.2012.10.018] [Medline: 23415116]
21. Pepper JK, Brewer NT. Electronic nicotine delivery system (electronic cigarette) awareness, use, reactions and beliefs: a systematic review. Tob Control 2014 Sep;23(5):375-384 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051122] [Medline: 24259045]
22. Kwon M, Park E. Perceptions and sentiments about electronic cigarettes on social media platforms: systematic review. JMIR Public Health Surveill 2020 Jan 15;6(1):e13673 [FREE Full text] [doi: 10.2196/13673] [Medline: 31939747]
23. McCausland K, Maycock B, Leaver T, Jancey J. The messages presented in electronic cigarette–related social media promotions and discussion: scoping review. J Med Internet Res 2019 Feb 05;21(2):e11953 [FREE Full text] [doi: 10.2196/11953] [Medline: 30720440]
24. Hua M, Alfi M, Talbot P. Health-related effects reported by electronic cigarette users in online forums. J Med Internet Res 2013 Apr 08;15(4):e59 [FREE Full text] [doi: 10.2196/jmir.2324] [Medline: 23567935]
25. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou M. "Okay, we get it. you vape": an analysis of geocoded content, context, and sentiment regarding e-cigarettes on Twitter. J Health Commun 2018 Jul 06;23(6):550-562. [doi: 10.1080/10810730.2018.1493057] [Medline: 29979920]
26. Unger JB, Escobedo P, Allem J, Soto DW, Chu K, Cruz T. Perceptions of secondhand e-cigarette aerosol among Twitter users. Tob Regul Sci 2016 Apr 01;2(2):146-152. [doi: 10.18001/trs.2.2.5]
27. Chen AT, Zhu S, Conway M. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. J Med Internet Res 2015 Sep 29;17(9):e220 [FREE Full text] [doi: 10.2196/jmir.4517] [Medline: 26420469]
28. Ghosh D, Guha R. What are we 'tweeting' about obesity? mapping tweets with topic modeling and geographic information system. Cartogr Geogr Inf Sci 2013 Mar;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]
29. National Cancer Prevention Month. American Association for Cancer Research. URL: https://www.aacr.org/patients-caregivers/awareness-months/national-cancer-prevention-month/ [accessed 2021-12-13]
30. Lung cancer risk factors. American Cancer Society. URL: https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/risk-factors.html [accessed 2021-12-13]
31. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]
32. Vraga EK, Stefanidis A, Lamprianidis G, Croitoru A, Crooks AT, Delamater PL, et al. Cancer and social media: a comparison of traffic about breast cancer, prostate cancer, and other reproductive cancers on Twitter and Instagram. J Health Commun 2018 Jan 09;23(2):181-189. [doi: 10.1080/10810730.2017.1421730] [Medline: 29313761]
33. Diddi P, Lundy LK. Organizational Twitter use: content analysis of Tweets during breast cancer awareness month. J Health Commun 2017 Mar 19;22(3):243-253. [doi: 10.1080/10810730.2016.1266716] [Medline: 28248621]
34. Lee EW, Ho SS, Chow JK, Wu YY, Yang Z. Communication and knowledge as motivators: understanding Singaporean women's perceived risks of breast cancer and intentions to engage in preventive measures. J Risk Res 2013 Aug;16(7):879-902. [doi: 10.1080/13669877.2012.761264]
35. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015 Oct 27;17(10):e243 [FREE Full text] [doi: 10.2196/jmir.4969] [Medline: 26508089]
36. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]

37. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for coding tobacco-related Twitter data: a systematic review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
38. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
39. Grün B, Hornik K. Topicmodels: an R package for fitting topic models. J Stat Soft 2011;40(13):1-30. [doi: 10.18637/jss.v040.i13]
40. Huang L, Ma J, Chen C. Topic detection from microblogs using T-LDA and perplexity. 2017 Presented at: 24th Asia-Pacific Software Engineering Conference Workshops (APSECW); December 4-8; Nanjing, Jiangsu, China p. 71-77. [doi: 10.1109/apsecw.2017.11]
41. Tobacco use – legislation. Centers for Disease Control and Prevention. URL: https://chronicdata.cdc.gov/browse?category=Legislation [accessed 2021-12-13]
42. American cancer society position on electronic cigarettes. American Cancer Society. URL: https://www.cancer.org/healthy/stay-away-from-tobacco/e-cigarettes-vaping/e-cigarette-position-statement.html [accessed 2021-12-13]
43. Cullen KA, Ambrose BK, Gentzke AS, Apelberg BJ, Jamal A, King BA. Use of electronic cigarettes and any tobacco product among middle and high school students - United States, 2011-2018. MMWR Morb Mortal Wkly Rep 2018 Nov 16;67(45):1276-1277 [FREE Full text] [doi: 10.15585/mmwr.mm6745a5] [Medline: 30439875]
44. Talking to youth and young adults to prevent e-cigarette use. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/features/ecigarettes-young-people/index.html [accessed 2021-12-13]
45. Levy DT, Sweanor D, Sanchez-Romero LM, O'Connor R, Goniewicz ML, Borland R. Altria-Juul Labs deal: why did it occur and what does it mean for the US nicotine delivery product market. Tob Control 2020 Dec 04;29(e1):e171-e174 [FREE Full text] [doi: 10.1136/tobaccocontrol-2019-055081] [Medline: 31484802]
46. Tan ASL, Soneji S, Moran MB, Choi K. JUUL Labs' sponsorship and the scientific integrity of vaping research. The Lancet 2019 Aug;394(10196):366-368. [doi: 10.1016/s0140-6736(19)31718-0]
47. Ackerman A, Etow A, Bartel S, Ribisl KM. Reducing the density and number of tobacco retailers: policy solutions and legal issues. Nicotine Tob Res 2017 Feb 28;19(2):133-140 [FREE Full text] [doi: 10.1093/ntr/ntw124] [Medline: 27127232]
48. Feuer W. Altria stock craters after $4.1 billion hit from Juul investment, blames growing legal risk. CNBC. URL: https://www.cnbc.com/2020/01/30/altria-is-taking-a-4point1-billion-charge-on-its-investment-in-juul.html [accessed 2021-12-13]
49. Tobacco use in Washington 2019. Truth Initiative. URL: https://truthinitiative.org/research-resources/smoking-region/tobacco-use-washington-2019 [accessed 2021-12-13]
50. Understanding youth vaping in Hawaii. The Hawaii Department of Health. URL: https://www.healthyhawaii.com/tobacco-free/vaping/ [accessed 2021-12-13]
51. Extinguishing the tobacco epidemic in Hawaii. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/about/osh/state-fact-sheets/hawaii/index.html [accessed 2021-12-13]

Abbreviations

CDC: Centers for Disease Control and Prevention
e-cigarette: electronic cigarette

Edited by R Kukafka; submitted 18.02.21; peer-reviewed by W Chu, C Hudak; comments to author 15.04.21; revised version received 08.06.21; accepted 08.11.21; published 29.12.21.

Please cite as:
Lu J, Lee EWJ
Examining Twitter Discourse on Electronic Cigarette and Tobacco Consumption During National Cancer Prevention Month in 2018: Topic Modeling and Geospatial Analysis
J Med Internet Res 2021;23(12):e28042
URL: https://www.jmir.org/2021/12/e28042
doi:10.2196/28042
PMID:34964716

Original Paper

Culture-Specific Observations in a Saudi Arabian Digital Home Health Care Program: Focus Group Discussions With Patients and Their Caregivers

Abdulaziz A Alodhayani[1], MD; Marwah Mazen Hassounah[1], MBBS, MPH; Fatima R Qadri[2,3], MSc; Noura A Abouammoh[1], MBBS, MPH, PhD; Zakiuddin Ahmed[2,4], MBBS; Abdullah M Aldahmash[5,6], PhD

[1]Department of Family and Community Medicine, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[2]Prince Naif Bin Abdulaziz Health Research Center, King Saud University, Riyadh, Saudi Arabia

[3]Medical Researches Company, iResearch, Riyadh, Saudi Arabia

[4]Riphah Institute of Healthcare Improvement & Safety, Riphah International University, Islamabad, Pakistan

[5]Stem Cell Unit, Department of Anatomy, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[6]College of Medicine, Vision Colleges, Riyadh, Saudi Arabia

Corresponding Author:
Fatima R Qadri, MSc
Medical Researches Company
iResearch
Ar Rawdah
Riyadh
Saudi Arabia
Phone: 966 547959808
Email: far.qadri18@gmail.com

Abstract

Background: There is growing evidence of the need to consider cultural factors in the design and implementation of digital health interventions. However, there is still inadequate knowledge pertaining to the aspects of the Saudi Arabian culture that need to be considered in the design and implementation of digital health programs, especially in the context of home health care services for patients who are chronically and terminally ill.

Objective: This study aims to explore the specific cultural factors related to patients and their caregivers from the perspective of physicians, nurses, and trainers that have influenced the pilot implementation of Remotely Accessible Healthcare At Home, a connected health program in the Home Health Care department at King Saud University Medical City, Riyadh, Saudi Arabia.

Methods: A qualitative study design was adopted to conduct a focus group discussion in July 2019 using a semistructured interview guide with 3 female and 4 male participants working as nurses, family physicians, and information technologists. Qualitative data obtained were analyzed using a thematic framework analysis.

Results: A total of 2 categories emerged from the focus group discussion that influenced the experiences of digital health program intervention: first, culture-related factors including language and communication, cultural views on using cameras during consultation, nonadherence to web-based consultations, and family role and commitment and second, caregiver characteristics in telemedicine that includes their skills and education and electronic literacy. Participants of this study revealed that indirect contact with patients and their family members may work as a barrier to proper communication through the Remotely Accessible Healthcare At Home program.

Conclusions: We recommend exploring the use of interpreters in digital health, creating awareness among the local population regarding privacy in digital health, and actively involving direct family members with the health care providers.

(J Med Internet Res 2021;23(12):e26002) doi:10.2196/26002

KEYWORDS

connected health; digital health; telehealth; telemedicine; culture; Islam; Arab; mobile phone

Introduction

Background

Digital health technology (including telemedicine, telehealth, remote patient monitoring, and connected health) that supports patient-centered care is a rapidly evolving field and is gaining popularity as a means of driving quality improvements and accessibility in health care [1]. Digital health is assumed to enhance the provision of health care services outside conventional hospital settings [2,3], thereby strengthening the development of home health care (HHC) services. Numerous benefits related to the implementation of digital health have been suggested, including improved quality of care, patient safety, accessibility, and cost minimization [4,5]. Their use is also linked with enhanced clinical outcomes, such as a decline in mortality and emergency room admission rates [6]. During the 2019 COVID-19 global pandemic, telemedicine and telehealth have provided opportunities for delivering health care remotely and have underpinned the beneficial role of digital health in the future [7].

Owing to these beneficial uses of technology in health care worldwide, there has been a dramatic growth in the number of digital health pilot programs [8]. However, a critical issue in this domain is that a considerable number of these pilot programs have been unsuccessful in progressing toward full-scale implementation and widespread adoption [9-11]. The reasons underlying the failure of these pilot programs have received substantial interest in the literature [12,13]. One of the key barriers identified in numerous studies pertains to cultural factors and influences [14-16]. Culture refers to “a system of knowledge, beliefs, patterns of behavior, artifacts, and institutions that are created, learned, and shared by a group of people” [17]. Culturally appropriate health-related interventions have been reported to improve clinical outcomes [18]. Likewise, cultural sensitivity is a key element for honoring patient-centered health care [19], and it has been identified as one of the significant factors needed to extend telemedicine projects to mainstream health care [20].

Being the cradle of the Islamic faith, Saudi Arabia is home to a predominantly Muslim population, and this shapes cultural and social life. In general, some characteristics of the Arab-Muslim culture that influence health include health beliefs, family relationships, health providers, diet and medications, life stages, views on death and dying, preventive health, gender issues, space, time, and communication [21]. Specific to Saudi Arabia, the language barrier between patients and providers has been well documented to adversely affect the quality of health care in Saudi Arabia [22,23].

The influence of culture on digital health practices in Saudi Arabia is the focus of a study by Kaliyadan et al [24], which included 166 patients and evaluated the application of a 4G smartphone for teledermatology. The study found that 14% of the mostly female patients refused photography of their skin lesions, citing social and cultural reasons [24]. In addition, a review article by Alkhalifah and Aldhalaan [25] that examined the use of telehealth services to support families with children diagnosed with autism spectrum disorder in rural areas of Saudi Arabia recognized that telehealth interventions developed in the West were culturally unsuitable in the local context, thus acting as a barrier to extensive adoption [25]. Hence, the provision of culturally appropriate digital health is crucial for delivering high-quality health care in Saudi Arabia.

Research Gap

There is still inadequate knowledge pertaining to the aspects of Saudi Arabian culture that need to be considered closely in the design and implementation of digital and connected health programs. This is especially relevant in the context of HHC services for vulnerable groups such as patients who are chronically and terminally ill that require longstanding treatment and management and often have complex unmet health care needs [26]. Information regarding the unique cultural factors of the patients and caregivers that influence the successful application and adoption of health technology will enable program managers and health care providers to design and deliver culturally sensitive digital health interventions for use in the local context [27]. This is especially vital in light of Saudi Arabia’s Vision 2030 National Transformation Program (2018-2020) that has reinforced the role of patient-centered telemedicine and digital health technology as a means of driving major developments in the health care sector and improving health care outreach [28]. Its vitality is also relevant in light of Saudi Arabia’s active digital health response to the COVID-19 pandemic [29].

Thus, through our study, we aim to explore the culture-specific observations relating to patients and their caregivers through the perspective of physicians, nurses, and trainers during the pilot implementation of a connected health program named Remotely Accessible Healthcare At Home (RAHAH) in the HHC department at King Saud University Medical City (KSUMC), Riyadh, Saudi Arabia.

Methods

Study Design

This is a phenomenological qualitative study using focus group discussions (FGDs) to explore experiences and opinions of health care workers (HCWs) and program trainers who have been using telehealth for HHC in the Saudi Arabian cultural context. FGD was used as the technique, and it reflects the process through which meaning was constructed collectively and can be regarded as naturalistic.

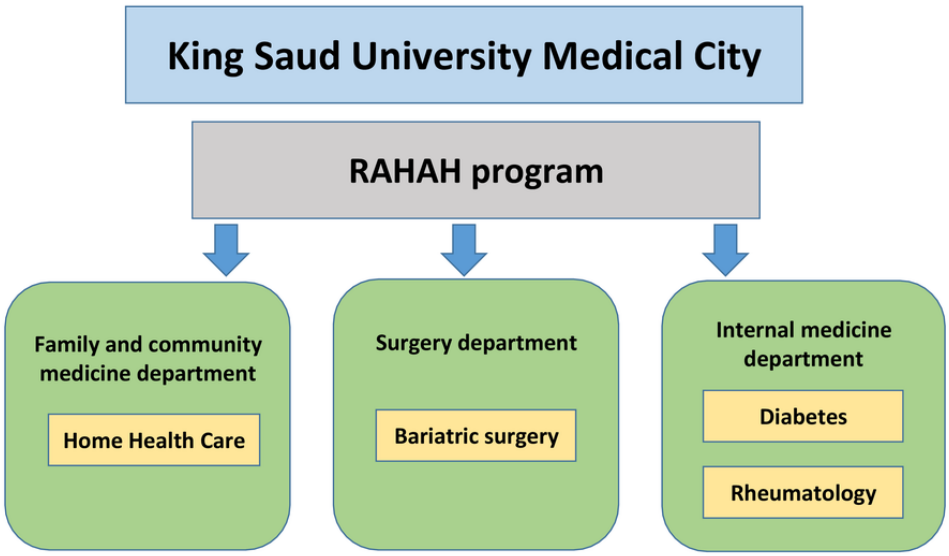
Some of the researchers (AAA, AMA, ZA, and FRQ) held an emic perspective as they worked in close contact with the RAHAH team. This assisted in bringing more knowledge, as they were more familiar with the expressions and ways of communicating and establishing trust between the researchers and participants. The others held an etic perspective (MMH and NAA), as they were only involved for the purpose of conducting the study. This allowed them to look at the participants experiences without a preconceived assumption.

Study Setting

KSUMC is a tertiary care hospital in the capital city of Riyadh. It is a leader in medical specialties, subspecialties, and

technologies. A connected health program called *RAHAH* was developed by Prince Naif Bin AbdulAziz Health Research Center at KSUMC. RAHAH uses a wide array of digital-based health tools on its platform, including patient health records, portals, mobile apps, remote monitoring devices (such as glucometers, blood pressure monitors, oximeters, and thermometers), teleconsultation, and e-response centers (through messaging and chat). In 2018, the HHC department at KSUMC embarked on piloting the RAHAH program in its department (Figure 1). This pilot program is being conducted in a phased approach, with geriatric patients constituting the first phase. There were 40 geriatric patients with chronic conditions such as diabetes mellitus and hypertension who were enrolled in this program when the study was conducted.

Figure 1. Different departments piloted by Remotely Accessible Healthcare At Home program including the Home Health Care department. RAHAH: Remotely Accessible Healthcare At Home.



Sampling and Recruitment

This research project was conducted at the HHC department, KSUMC where 6 HCWs were dedicated to piloting the RAHAH program. RAHAH provides technical liaisons to different departments at KSUMC to implement the RAHAH program. These liaisons also train patients, caregivers, physicians, nurses, and allied health staff on how to use the service, observe and curate feedback from users, and help in cases of technical hurdles. They provide valuable insight into the audience of interest (the HHC team and patients or caregivers) because of their close and frequent contact with them. For the HHC department at the time of this study, there were 2 RAHAH technical liaisons. This project's research team comprises administrative researcher physician figures from the HHC department and RAHAH program (AAA, AMA, and ZA), 2 qualitative research public health physicians from community medicine units (MMH and NAA), and a RAHAH researcher (FRQ).

The criteria for participant selection included those working on the RAHAH project in the HHC department at KSUMC. Those who were not directly involved in the HHC department at KSUMC were excluded. The total number of team members working on the RAHAH pilot in the HHC department was 8 (6/8, 75% HCWs and 2/8, 25% technical liaisons). They were all invited by researchers MMH and FRQ (who are not associated with HHC) via email to participate in the focus group. It was explained to all participants that their decision to participate would not have any implication on their jobs. An invitee declined (1 of the 2 RAHAH technical liaisons), and the remainder accepted by responding to the email. Thus, a total of 7 participants attended the FGD. Reminders for the invitation date and time were sent out via SMS text messages and emails with no prior knowledge of the actual number that will be attending the focus group. The focus group time was voided from the HHC team schedule to provide an opportunity to those who wanted to participate in the research.

Data Collection

The discussion guide was developed in English based on the study objectives, previous studies [15,21,30-33], professionals' experiences [34], and personal familiarity and clinical experiences with the Saudi Arabian culture (Multimedia Appendix 1). It followed a semistructured format with open-ended questions that facilitated the development of emergent themes during the discussion.

The discussion guide was tested with 1 RAHAH trainer who was not invited to the FGD. A researcher tested the guide for ease of understanding and checked whether the answers met the requirements of the question. Consequently, some of the guide questions were rephrased, and the definition of *culture* was added to the introduction.

The FGD was conducted on July 28, 2019, for 2 hours by 2 researchers, MMH and FRQ. MMH led the discussion, as she is bilingual in Arabic and English, whereas FRQ took notes of the discussion and kept track of the time. The discussion was conducted in a quiet meeting room at KSUMC on July 28, 2019, and lasted for an hour.

The discussion was in English, which is the main language spoken by workers in KSUMC. However, to enhance the credibility of the findings, participants were given the option

to use Arabic. For each question, the focus group facilitator gave ample time for all participants to share their thoughts and at times used prompters to further explore the answers. The facilitator would move on to the next question when she reached saturation with the answers.

The discussion was audiotaped and transcribed verbatim by a transcription service. The transcript had time stamps on speaker change, and MMH and FRQ filled in unintelligible and Arabic audio, deidentified the participants and names, crosschecked the text with audio, and refined and finalized the transcript for analysis.

Having different perspectives among the authors, using probing questions, using ≥1 coder, and having transparency in reporting the results added to the trustworthiness of the study. A focus group method was chosen as it is the best technique to elicit shared experiences, permitting participants to raise topics that they deem to be important and allowing people to probe each other's reasons for holding certain views. All of the above enriched the data and supported the credibility of the study.

Data Analysis

We conducted a thematic framework analysis as described by Ritchie and Spencer [35]. A holistic overview of the data set and familiarization with the range, depth, and diversity of the participants' responses was performed by NAA and MMH. Similar codes were developed by both authors, and an agreement on an initial coding frame was achieved with the flexibility to enable other codes to be added and discussed regularly. Key themes were developed, and quotations were used to support the provided evidence. NVivo software (version 11.4.2; QSR International) [36] was used to manage the data.

Ethical Considerations

Ethical approval was obtained from the institutional review board of King Saud University College of Medicine in June 2019 (project number E-18-3914) based on the Declaration of Helsinki. Participation in the study was voluntary, maintaining the right of the participants to withdraw at any time. The participants were fully informed regarding the study aims, objectives, methods, and how the interview material would be recorded and protected. All the information provided by the participants was kept confidential, the participants' names were coded, and possible identifiers were omitted from the final transcript. As the head of the HHC department and RAHAH program managers are researchers in this study, they were not involved in the facilitation or transcription of this focus group to avoid the power effect of the relationship between managers and employees to influence the latter's responses.

Results

Overview

A total of 7 participants were included in the focus group (Table 1), and 2 categories emerged from the FGD. These were (1) culture-related factors, including language and communication, cultural views on using cameras during consultation, nonadherence to web-based consultations, and family role and commitment and (2) caregiver characteristics in telemedicine, including their skills, education, and electronic literacy (Textbox 1).

Table 1. Demographic data of the study participants.

Participant number	Nationality	Gender	Time working in home care and telemedicine	Specialty
1	Saudi Arabia	Female	3 years	Information technology
2	South Asia	Male	3 years	Family physician
3	South Asia	Male	1 year	Family physician
4	Southeast Asia	Male	4 years	Nurse
5	Southeast Asia	Male	3 years	Nurse
6	South Asia	Female	3 years	Nurse
7	Saudi Arabia	Female	3 months	Nurse

Textbox 1. Emerging themes in the experiences of health care workers.

Culture-related factors

- Language and communication
- Communicating with the caregiver
- Behavior and media used in urgent communication
- Bilingualism
- Cultural views on using the camera
- Nonadherence to web-based consultations
- Family role and commitment

Caregiver characteristics

- Skills and education
- Electronic literacy

Culture-Related Factors

A total of 4 themes emerged from this category: language and communication, cultural views on using the camera, nonadherence to web-based consultations, and family role and commitment.

Language and Communication

Communicating With the Caregiver

The participants acknowledged that dealing with the caregivers rather than directly with patients is inevitable, as almost all patients who are registered in the RAHAH program at the HHC department are diagnosed with cognitive impairment. This is evident in the response of participant 2:

> *Most of the patients are not able to express themselves...they are demented, and they cannot express themselves.*

As family members are usually busy during the day, they assign a housemaid or sometimes a nurse to escort the patient. A participant explained the following: "We are communicating with the attendants who are hired ones, not the real relatives of the patient" [participant 2].

Another participant added the following: "Very few, uh, family members are communicating with us" [participant 1].

Caregivers were seen by the participants as an essential link between the patients and the health care team. According to them, they train all caregivers to assist in patient care: "We train them on feeding, giving medications, vital signs recording" [participant 2].

They correspondingly acknowledged that if the caregiver was sincere, good patient outcomes could be observed. A participant explained as follows:

> *Dedication is very important for the caregiver because we have some patients with bed sores...So, if she is dedicated [you can see] good progress and good healing of the wound...But if they are not dedicated and they are not concerned, it will linger on.* [participant 3]

Furthermore, discrepancies in the information shared by patients and caregivers were noted. A participant explained the following:

> *Even when their caregivers told us that she [the patient] was having pain in her knees, she was not eating, she—didn't go to the toilet for like three, four days and I asked her, "Mama, how are you? Are you okay?" "Yeah, everything is good."* [participant 7]

Moreover, handover of care between caregivers and family members was described as *poor,* as they do not involve the health care team in the process of care transition to family members. A participant stated the following: "They [caregivers] will not even tell us when they are travelling" [participant 7].

Behavior and Media Used in Urgent Communication

According to the participants, phone calls to their offices or messages on WhatsApp were the modes of communication in case of emergencies. A participant explained the following:

> *We are not watching RAHAH all the time...They [patients] call our office. If there is some urgent problem...they send the pictures, messages on WhatsApp and immediately we connect and then call them and see what the problem is.* [participant 2]

Participants felt comfortable providing their personal cell phone contacts, as they believed that some patients might need immediate attention:

> *If the nasogastric tube is out...So, they communicate with us because our patients need to be fed...or if the catheter is pulled out, they need to be relieved.* [participant 3]

A participant expressed his inconvenience in dealing remotely with some caregivers in the case of an emergency. He believed that caregivers' expressed stress could be exaggerated, and the situation could be misinterpreted by the HCW:

> *You directly see they are very hysterical, especially ladies, and shouting "the patient is dying! The patient is dying!" or "there's blood coming out" or "please, we cannot give medicine anymore." So, once they are*

> *hysterical...They do create a hysterical situation for the treatment team.* [participant 2]

Bilingualism

The participants denied language as a barrier to communication. They noticed that their caregivers or relatives spoke both Arabic and English, which made it easier for them to communicate. Participant 3 explained it as follows: "If you go back years and years, most people were speaking Arabic, but now I have noticed most of the people there speak English fluently too."

In contrast, another participant noted the following:

> *Sometimes they do not know how to use the system because the RAHAH system is an English based system...Others [caregivers] from African countries do not speak English, so they do not communicate with us directly and they just work under the supervision of some ladies of the house.* [participant 2]

The problem of caregivers with limited knowledge of the English language contributes to difficulties in communication with health care providers and eventually influences the level of care provided to the patient. Those who are fluent in English would be able to operate the telehealth program effectively and understand the treatment team instructions better compared with those who do not speak English:

> *Some female foreigner caregivers can speak English and they can operate the telehealth system. They are the most helpful to us, taking pictures, and sending us messages.* [participant 2]

> *Others are housemaids who are not trained in the English language and do not communicate well with us; they just work under the supervision of family members. They are trained to clean the patient's bed, bedsheets, and things, and they are not trained to take care of medical problems. They are also not able to use the telemedicine system.* [participant 2]

Cultural Views on Using the Camera

Participants commented on the acceptance of camera use by patients during web-based consultations. A participant noted the following:

> *Even if I ask them to show me the patient, they are reluctant. They say "I will tell you everything instead."* [participant 2]

Some participants believed that this is rooted in a religious viewpoint, acknowledging that Saudi Arabian culture is shaped by the Islamic religious background. A participant remarked the following: "Some might think that showing themselves on camera is Haram [religiously prohibited]" [participant 1].

Participants also speculated that most female patients were uncomfortable showing themselves on camera, as they covered their faces from male strangers. A participant further opined that the patients or caregivers feared using the camera and technology during communication with their health care provider and expressed concerns regarding their privacy:

> *They do not want to use the camera...because it could be recorded. They are afraid that technology is so advanced, and cameras can record their (face)...stalking...taking photos.* [participant 2]

The participants showed full acceptance of the patients' preferences. Moreover, most of them preferred not to ask female patients to show themselves on camera anymore, as they knew this might not be acceptable to them and only tried to see parts of the body when necessary. Male patients, on the other hand, do not generally have reservations in this regard. A participant noted the following:

> *We do not request any exposure on camera. For females, we do not request to examine them on the telemedicine system...or to show parts of their body unless absolutely necessary. But in males...it is okay.* [participant 3]

Nonadherence to Web-Based Consultations

Participants perceived that patients' attitudes toward attending their web-based appointments were different from face-to-face hospital appointments. Some patients do not take their web-based appointments seriously. A participant explained the following:

> *They are not particular about this appointment and whether to attend it or not...In out-patient clinics, people are careful because if they do not show up then they will not get another appointment for three months. Here, there is no such problem. They can get another one soon.* [participant 2]

Family Role and Commitment

Most participants acknowledged the key role of family members in facilitating the implementation of the RAHAH program in HHC and communicating with health care providers. Moreover, their role in reaching an optimum state of health care outcomes was highlighted. The participants mentioned the following:

> *The main role is the family's.* [participant 3]

> *If the family is proactive, we can expect that the patient has no acute conditions...their health status is optimum and meet expectations.* [participant 5]

However, the participants conveyed that family members were unavailable, most of the time, to take care of the patients:

> *...because family members are working outside the home, and they are busy.* [participant 3]

> *Men generally are not at home...Even ladies are working, such as teachers.* [participant 2]

Consequently, physicians generally rely on a hired nonnational caregiver to provide them with health-related information about the patient, and sometimes, the team may find it difficult to communicate with family members when the caregiver is unavailable. A participant noted the following:

> *They [the family member] will be updated by the caregiver, usually a housemaid. And whenever we ask more questions, they need to call the caregiver for more clarification. This may create a problem with telling the truth.* [participant 7]

The participants further acknowledged that the commitment of families varied regarding their involvement in patient care. A participant reported a few incidents that highlighted the lack of commitment by some family members:

> *Do you remember what happened last week with us? The patient who had not taken medicine...he used to take medicine at 9 or 10 AM. Why? Because the caregiver was traveling. The lady of the house was there at home, and she said, "I don't know how to give the medications. I told the caregiver that before she leaves, she should hand over the care."* [participant 3]

A participant reported that some families hindered the successful implementation of the telehealth program and recounted an instance in which the family did not provide the Wi-Fi password to the caregiver in their absence:

> *I have encountered two caregivers who want to use RAHAH, but the problem is they do not have Wi-Fi. I mean the family do not give the password to the caregivers.* [participant 4]

In short, the quality of the telehealth experience is affected by specific culture-related factors, including language and mode of communication, religious views of the patients or caregivers, and family role and commitment to patients' care.

Caregiver Characteristics

As caregivers are the main link between patients and health care providers, under this category, caregiver characteristics that were commonly observed by the participants were classified into the following themes: skills and education and electronic literacy of the caregiver.

Skills and Education

The participants highlighted several individual skills of the caregivers that influenced the care provided to patients. For example, a participant noted the following:

> *We see all types and all categories of caregivers. Some are trained nurses, some are nursing assistants, and some are only housemaids.* [participant 3]

Educated caregivers prove to be valuable in driving an effective telehealth consultation, as they are able to identify dangerous signs and symptoms in the patient and inform the treating team in a timely manner. A participant described incidents where an educated caregiver was able to use the telehealth program to connect with the health care team and alert them regarding unusual symptoms such as blood in the stool and discoloration of the patient's leg, which subsequently saved the life of the patient:

> *So, an educated attendant is the main person who is helping us in telemedicine...remember that lady? She sent out the photos of the blood in the stool. Therefore, we immediately contacted her and then we went there to see the patient, and found out that he had colon cancer...There was a blood clot in the leg that we recognized from picture that she sent.* [participant 2]

Electronic Literacy

According to the participants, some caregivers had limited electronic literacy and found it difficult to deal with the technology necessary to conduct a telehealth consultation:

> *Yeah, um, few times I faced difficulties in leading the caregiver to just sign up to Gmail...even the application...sometimes they have difficulties with that.* [participant 1]

In addition, some patients and caregivers showed low self-confidence in using the technology. A participant explained the following:

> *One patient said "I do not know how to use the application. I do not know how to use technology. I am not good with phones."* [participant 7]

Discussion

Principal Findings

The RAHAH program applies a user-centered design approach to its development process. The program team is in constant communication with the health care providers piloting it, as well as patients and caregivers receiving care through the program. The program team also conducts frequent surveys and observation visits to better understand the user experience and enhance it. This qualitative study aligns with this approach of adopting the bottom-up design of the program for better acceptability, usability, and satisfaction. The pathway of communication in the RAHAH program is depicted in Figure 2, where the HCWs communicate with the caregiver, as most patients have difficulties in verbal communication, who in turn communicate with the family members. HCWs communicate with family members only when needed or in emergency cases.

Our findings show that HCWs communicate directly with caregivers who are characterized by some features that may hinder proper telemedicine-based communication and hence affect the quality of care provided by the health care team. This finding aligns well with previous studies that suggested cultural influences impede the full-fledged implementation and adoption of digital health systems [10-12]. Moreover, although HCWs using telemedicine believe that better health outcomes are associated with family involvement, their direct communication with the family members of the patients is lacking. This can be explained by the changing social culture in Saudi Arabia in terms of more women joining the workforce [37] and the breakdown of the extended family structure [38].

Although language was not explicitly reflected as a significant barrier by the participants, this aspect must be considered because the probability of challenges related to language and communication is high in Saudi Arabia, as many of the caregivers are expatriates and do not speak Arabic [22,23]. This is a matter of concern, as poor patient–provider communication is linked to poor diagnosis, treatment, and medication instructions and a significantly higher risk of serious medical events [39-43]. Bilingual providers are preferred to bridge the patient–provider gap. In this regard, researchers suggest that professionally trained interpreters can provide a high-quality, culturally competent language in the absence of the services of

bilingual clinicians [44]. However, a scarcity of interpreters has been reported previously [45], and in Saudi Arabia, a systematic review suggested that the cultural and language training currently provided to expatriate nurses is not fulfilling its purpose [46].

The findings of this study are consistent with earlier studies that highlight the cultural beliefs affecting camera use in teleconsultations, especially among women [24,47-49]. The culture-religious belief that inhibits women from showing themselves on camera is deeply rooted in decades-old *Fatwas* (scholarly religious decisions) that forbade photography in all its forms and the revealing of any part of a female's body to men, even in medical situations. These *Fatwas* have been amended to accommodate the modern-day context but remain a deterrent in the consciousness of some patients. It was clearly important for the participants of this study to be culturally sensitive when delivering their services using RAHAH. However, they may hinder a successful telehealth experience by refraining from asking female patients to appear on camera or to show some body parts.

Figure 2. Pathway of health communication using telemedicine in Remotely Accessible Healthcare At Home program in the Home Health Care department.

Studies have largely reported technology-learning barriers in telemedicine [50-52]. Our findings contrast with previously published studies in which a positive attitude toward telemedicine was evident in patients with cancer [53], as well as in those with lung diseases [54]. In their large survey study, Edwards et al [55] found that patients with depression and those with a high risk of cerebrovascular disease have a moderate interest in phone-, email-, and internet-based services, whereas interest in social media–based services was lower. In Saudi Arabia, it was found that several barriers hinder decision-makers of health care facilities from adopting and implementing telemedicine in their health care facilities. These barriers were the availability of adequate, sustainable funding and financial support for operations; the reimbursement for telemedicine services; the quality of information; the cost-effectiveness of telemedicine; and cultural and social constraints [56]. Another local study highlighted factors such as privacy, equipment cost, lack of training, information, and communication technology issues as significant barriers to the adoption of telemedicine in hospitals [57]. However, the latter studies focused on patients or HCWs mediating the consultation in the absence of caregivers.

Regarding the recurring *no show* behavior, it could be curbed with punitive procedures and communicated to caregivers when onboarding the program. The cost-effectiveness of the implemented telehealth program should be assessed with regard to nonadherence to web-based consultations [58]. The *no show* behavior in the served population in this study could be explained by proximity to the hospital premises. In this case, telehealth was not the only option for patients. Telehealth services for individuals living in rural areas have observed different behavior [25,59].

Overall, it is important to empower patients and caregivers in the context of digital health care. Our study can be looked at as a piece of a whole. The whole picture includes patients, caregivers, and HCWs. The findings of this study can be compared against perspectives on the experience of the primary audience of the RAHAH program later on to improve the overall communication process.

Study Limitations

Potential information recall bias among participants must be considered in similar studies in the future. This study would have benefited from triangulation with the primary audience (patients and caregivers). In this study, patients were not interviewed because of their health conditions, whereas caregivers spoke different languages, which required unavailable resources for translation to conduct the interviews and transcribe them. Future projects could triangulate with other departments' findings and RAHAH's user analytics and should capture the experiences of patients or their family members. The study team dedicated effort to concealing the participants' identities by not involving research team members who are also HHC or RAHAH administrative figures in the focus group or analysis phases.

Conclusions

The study results illustrate the importance of improving awareness among the public and health care teams to reassure the patients and caregivers that their privacy and preferences with respect to the socioreligious culture are vital. HCWs should remind caregivers of the importance of following a proper handover of care, should the need arise. Furthermore, because caregivers are not always professionally trained in nursing and personal characteristics affect their capabilities, it is crucial that HCWs communicate directly with family members, providing a summary of the teleconsultation conducted in their absence. More research could focus on the feasibility of using interpreters to improve the care provided via telehealth.

Acknowledgments

This study was supported by the Deanship of Scientific Research, Research Chair for Evidence-Based Healthcare and Knowledge Translation, King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group discussion guide.
[DOCX File , 20 KB - jmir_v23i12e26002_app1.docx]

References

1. Chu L, Shah AG, Rouholiman D, Riggare S, Gamble JG. Patient-centric strategies in digital health. In: Rivas H, Wac K, editors. Digital Health. Health Informatics. Cham: Springer; Jan 03, 2018.
2. Voran D. Telemedicine and beyond. Mo Med 2015;112(2):129-135 [FREE Full text] [Medline: 25958658]
3. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
4. Agboola S, Kvedar J. Telemedicine and patient safety. AHRQ Patient Safety Network. 2016. URL: https://psnet.ahrq.gov/perspective/telemedicine-and-patient-safety [accessed 2021-11-18]
5. Di Cerbo A, Morales-Medina JC, Palmieri B, Iannitti T. Narrative review of telemedicine consultation in medical practice. Patient Prefer Adherence 2015;9:65-75 [FREE Full text] [doi: 10.2147/PPA.S61617] [Medline: 25609928]
6. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012 Jun 21;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874] [Medline: 22723612]
7. 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]
8. Colaci D, Chaudhri S, Vasan A. mHealth interventions in low-income countries to address maternal health: a systematic review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
9. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
10. McCann D. A Ugandan mHealth moratorium is a good thing. ICTworks. 2012. URL: https://www.ictworks.org/ugandan-mhealth-moratorium-good-thing/#.W_vPUoczbIU [accessed 2021-11-18]
11. National eHealth Strategy Toolkit. International Telecommunication Union. 2012. URL: https://apps.who.int/iris/handle/10665/75211 [accessed 2021-11-18]
12. Luna D, Almerares A, Mayan 3rd JC, González Bernaldo de Quirós F, Otero C. Health informatics in developing countries: going beyond pilot practices to sustainable implementations: a review of the current challenges. Healthc Inform Res 2014 Jan;20(1):3-10 [FREE Full text] [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]
13. Zanaboni P, Wootton R. Adoption of telemedicine: from pilot stage to routine delivery. BMC Med Inform Decis Mak 2012 Jan 04;12:1 [FREE Full text] [doi: 10.1186/1472-6947-12-1] [Medline: 22217121]
14. Ly BA, Labonté R, Bourgeault IL, Niang MN. The individual and contextual determinants of the use of telemedicine: a descriptive study of the perceptions of Senegal's physicians and telemedicine projects managers. PLoS One 2017;12(7):e0181070 [FREE Full text] [doi: 10.1371/journal.pone.0181070] [Medline: 28732028]
15. Sundin P, Callan J, Mehta K. Why do entrepreneurial mHealth ventures in the developing world fail to scale? J Med Eng Technol 2016;40(7-8):444-457. [doi: 10.1080/03091902.2016.1213901] [Medline: 27686003]
16. Chung J, Thompson HJ, Joe J, Hall A, Demiris G. Examining Korean and Korean American older adults' perceived acceptability of home-based monitoring technologies in the context of culture. Inform Health Soc Care 2017 Jan;42(1):61-76. [doi: 10.3109/17538157.2016.1160244] [Medline: 27100664]
17. Guest K. Cultural Anthropology: A Toolkit for a Global Age. New York: W. W. Norton and Company; 2014.
18. Zeh P, Sandhu HK, Cannaby AM, Sturt JA. The impact of culturally competent diabetes care interventions for improving diabetes-related outcomes in ethnic minority groups: a systematic review. Diabet Med 2012 Oct;29(10):1237-1252. [doi: 10.1111/j.1464-5491.2012.03701.x] [Medline: 22553954]
19. Tucker CM, Marsiske M, Rice KG, Nielson JJ, Herman K. Patient-centered culturally sensitive health care: model testing and refinement. Health Psychol 2011 May;30(3):342-350 [FREE Full text] [doi: 10.1037/a0022967] [Medline: 21553978]
20. Eighteen critical success factors for deploying telemedicine. Momentum. 2014. URL: http://www.telemedicine-momentum.eu/wp [accessed 2021-11-28]
21. Haddad LG, Hoeman SP. Home healthcare and the Arab-American client. Home Healthc Nurse 2000 Mar;18(3):189-197. [doi: 10.1097/00004045-200003000-00011] [Medline: 11951245]
22. Abouammoh NA, Barnes S, Goyder E. Providing lifestyle advice to people with type 2 diabetes from different cultures: a qualitative investigation. Prim Care Diabetes 2016;18(6):283-287.
23. Alshammari M, Duff J, Guilhermino M. Barriers to nurse-patient communication in Saudi Arabia: an integrative review. BMC Nurs 2019 Dec 03;18:61 [FREE Full text] [doi: 10.1186/s12912-019-0385-4] [Medline: 31827387]

24. Kaliyadan F, Amin TT, Kuruvilla J, Ali WH. Mobile teledermatology–patient satisfaction, diagnostic and management concordance, and factors affecting patient refusal to participate in Saudi Arabia. J Telemed Telecare 2013 Sep;19(6):315-319. [doi: 10.1177/1357633X13501778] [Medline: 24163295]
25. Alkhalifah S, Aldhalaan H. Telehealth services for children with autism spectrum disorders in rural areas of the Kingdom of Saudi Arabia: overview and recommendations. JMIR Pediatr Parent 2018 Nov 15;1(2):e11402 [FREE Full text] [doi: 10.2196/11402] [Medline: 31518306]
26. Herr M, Arvieu J, Aegerter P, Robine J, Ankri J. Unmet health care needs of older people: prevalence and predictors in a French cross-sectional survey. Eur J Public Health 2014 Oct;24(5):808-813 [FREE Full text] [doi: 10.1093/eurpub/ckt179] [Medline: 24287029]
27. Aldahmash AM, Ahmed Z, Qadri FR, Thapa S, AlMuammar AM. Implementing a connected health intervention for remote patient monitoring in Saudi Arabia and Pakistan: explaining 'the what' and 'the how'. Global Health 2019 Mar 08;15(1):20 [FREE Full text] [doi: 10.1186/s12992-019-0462-1] [Medline: 30850012]
28. National Transformation Program Delivery Plan 2018-2020. Vision 2030. URL: https://vision2030.gov.sa/en/ntp [accessed 2021-11-28]
29. Hassounah M, Raheel H, Alhefzi M. Digital response during the COVID-19 pandemic in Saudi Arabia. J Med Internet Res 2020 Sep 01;22(9):e19338 [FREE Full text] [doi: 10.2196/19338] [Medline: 32790642]
30. Erwin DO, Treviño M, Saad-Harfouche FG, Rodriguez EM, Gage E, Jandorf L. Contextualizing diversity and culture within cancer control interventions for Latinas: changing interventions, not cultures. Soc Sci Med 2010 Aug;71(4):693-701. [doi: 10.1016/j.socscimed.2010.05.005] [Medline: 20646810]
31. Mendieta M, Buckingham R, Kietzman J, Helal A, Parker S. Understanding the barriers to hospice care in Saudi Arabia. J Palliat Med 2015 Nov;18(11):912. [doi: 10.1089/jpm.2015.0259] [Medline: 26406930]
32. Hilty D, Evangelatos G, Valasquez G, Le C, Sosa J. Telehealth for rural diverse populations: cultural and telebehavioral competencies and practical approaches for clinical services. J Technol Behav Sci 2018 Apr 3;3(3):206-220 [FREE Full text] [doi: 10.1007/s41347-018-0054-6]
33. Shore JH, Savin DM, Novins D, Manson SM. Cultural aspects of telepsychiatry. J Telemed Telecare 2006;12(3):116-121. [doi: 10.1258/135763306776738602] [Medline: 16638232]
34. Alotaibi HK. Cultural considerations in patient experience. In: Proceedings of the 2nd International Patient Experience Summit. 2019 Presented at: 2nd International Patient Experience Summit; Feb 4-6, 2019; Riyadh, Saudi Arabia.
35. Wordpress. URL: https://tinyurl.com/3nkp45ss [accessed 2021-12-01]
36. Unlock insights in your data with powerful analysis. NVIVO. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-11-18]
37. Saudi women rising up in business in line with Vision 2030. World Bank Group. 2020. URL: https://www.worldbank.org/en/news/feature/2020/03/11/saudi-women-rising-up-in-business-in-line-with-vision-2030 [accessed 2021-11-18]
38. Almalki S, Ganong L. Family life education in Saudi Arabia. In: Robila M, Taylor AC, editors. Perspectives on Family Life Education. Cham: Springer International Publishing; 2018.
39. Al Shamsi H, Almutairi A, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. Oman Med J 2020 Apr 30;35(2):e122 [FREE Full text] [doi: 10.5001/omj.2020.40] [Medline: 32411417]
40. Bonacruz Kazzi G, Cooper C. Barriers to the use of interpreters in emergency room paediatric consultations. J Paediatr Child Health 2003;39(4):259-263. [doi: 10.1046/j.1440-1754.2003.00135.x] [Medline: 12755930]
41. Lasater LM, Davidson AJ, Steiner JF, Mehler PS. Glycemic control in English- vs Spanish-speaking Hispanic patients with type 2 diabetes mellitus. Arch Intern Med 2001 Jan 08;161(1):77-82. [doi: 10.1001/archinte.161.1.77] [Medline: 11146701]
42. Shapiro J, Saltzer E. Cross-cultural aspects of physician-patient communications patterns. Urban Health 1981 Dec;10(10):10-15. [Medline: 10254342]
43. Cohen AL, Rivara F, Marcuse EK, McPhillips H, Davis R. Are language barriers associated with serious medical events in hospitalized pediatric patients? Pediatrics 2005 Sep;116(3):575-579. [doi: 10.1542/peds.2005-0521] [Medline: 16140695]
44. Juckett G, Unger K. Appropriate use of medical interpreters. Am Fam Physician 2014 Oct 01;90(7):476-480 [FREE Full text] [Medline: 25369625]
45. Jaeger FN, Pellaud N, Laville B, Klauser P. Barriers to and solutions for addressing insufficient professional interpreter use in primary healthcare. BMC Health Serv Res 2019 Oct 25;19(1):753 [FREE Full text] [doi: 10.1186/s12913-019-4628-6] [Medline: 31653211]
46. Alshammari M, Duff J, Guilhermino M. Barriers to nurse-patient communication in Saudi Arabia: an integrative review. BMC Nurs 2019;18:61 [FREE Full text] [doi: 10.1186/s12912-019-0385-4] [Medline: 31827387]
47. Kuziemsky CE, Gogia SB, Househ M, Petersen C, Basu A. Balancing health information exchange and privacy governance from a patient-centred connected health and telehealth perspective. Yearb Med Inform 2018 Aug;27(1):48-54 [FREE Full text] [doi: 10.1055/s-0038-1641195] [Medline: 29681043]
48. Alshammari F, Hassan S. Perceptions, preferences and experiences of telemedicine among users of information and communication technology in Saudi Arabia. J Health Inform Dev Ctries 2019;13(1) [FREE Full text]
49. Padela AI, Rodriguez del Pozo P. Muslim patients and cross-gender interactions in medicine: an Islamic bioethical perspective. J Med Ethics 2011 Jan;37(1):40-44. [doi: 10.1136/jme.2010.037614] [Medline: 21041237]

50. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012 Jul 26;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]
51. El-Mahalli AA, El-Khafif SH, Al-Qahtani MF. Successes and challenges in the implementation and application of telemedicine in the eastern province of Saudi Arabia. Perspect Health Inf Manag 2012;9:1-27 [FREE Full text] [Medline: 23209455]
52. Uscher-Pines L, Kahn JM. Barriers and facilitators to pediatric emergency telemedicine in the United States. Telemed J E Health 2014 Nov;20(11):990-996 [FREE Full text] [doi: 10.1089/tmj.2014.0015] [Medline: 25238565]
53. Girault A, Ferrua M, Lalloué B, Sicotte C, Fourcade A, Yatim F, et al. Internet-based technologies to improve cancer care coordination: current use and attitudes among cancer patients. Eur J Cancer 2015 Mar;51(4):551-557. [doi: 10.1016/j.ejca.2014.12.001] [Medline: 25661828]
54. Hofstede J, de Bie J, van Wijngaarden B, Heijmans M. Knowledge, use and attitude toward eHealth among patients with chronic lung diseases. Int J Med Inform 2014 Dec;83(12):967-974. [doi: 10.1016/j.ijmedinf.2014.08.011] [Medline: 25269992]
55. 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 08;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
56. Alaboudi A, Atkins A, Sharp B, Balkhair A, Alzahrani M, Sunbul T. Barriers and challenges in adopting Saudi telemedicine network: the perceptions of decision makers of healthcare facilities in Saudi Arabia. J Infect Public Health 2016;9(6):725-733 [FREE Full text] [doi: 10.1016/j.jiph.2016.09.001] [Medline: 27649882]
57. Albarrak AI, Mohammed R, Almarshoud N, Almujalli L, Aljaeed R, Altuwaijiri S, et al. Assessment of physician's knowledge, perception and willingness of telemedicine in Riyadh region, Saudi Arabia. J Infect Public Health 2021 Jan;14(1):97-102 [FREE Full text] [doi: 10.1016/j.jiph.2019.04.006] [Medline: 31060975]
58. Alghamdi SM, Janaudis-Ferreira T, Alhasani R, Ahmed S. Acceptance, adherence and dropout rates of individuals with COPD approached in telehealth interventions: a protocol for systematic review and meta-analysis. BMJ Open 2019 Apr 25;9(4):e026794 [FREE Full text] [doi: 10.1136/bmjopen-2018-026794] [Medline: 31028042]
59. Ahmed M, Ahmed S, Osman M. Telemedicine and teleradiology in Saudi Arabia. IOSR J Dent Med Sci 2014;13(2):88-90 [FREE Full text] [doi: 10.9790/0853-13228890]

Abbreviations

FGD: focus group discussion
HCW: health care worker
HHC: home health care
KSUMC: King Saud University Medical City
RAHAH: Remotely Accessible Healthcare At Home

Edited by R Kukafka; submitted 25.11.20; peer-reviewed by E van der Velde, R Aldekhyyel, H Mehdizadeh, S Alghamdi; comments to author 15.03.21; revised version received 09.05.21; accepted 18.10.21; published 08.12.21.

Please cite as:
Alodhayani AA, Hassounah MM, Qadri FR, Abouammoh NA, Ahmed Z, Aldahmash AM
Culture-Specific Observations in a Saudi Arabian Digital Home Health Care Program: Focus Group Discussions With Patients and Their Caregivers
J Med Internet Res 2021;23(12):e26002
URL: https://www.jmir.org/2021/12/e26002
doi:10.2196/26002
PMID:34889740

Review

Implementation, Adoption, and Perceptions of Telemental Health During the COVID-19 Pandemic: Systematic Review

Rebecca Appleton[1*], PhD; Julie Williams[2*], PhD; Norha Vera San Juan[3*], PhD; Justin J Needle[4], PhD; Merle Schlief[1], MSc; Harriet Jordan[2], MSc; Luke Sheridan Rains[1], PhD; Lucy Goulding[5], PhD; Monika Badhan[6], MBBS, MRCPsych; Emily Roxburgh[7], MSc; Phoebe Barnett[1,8], MSc; Spyros Spyridonidis[1], MSc; Magdalena Tomaskova[1], MA; Jiping Mo[1], MSc; Jasmine Harju-Seppänen[9], MSc; Zoë Haime[9], MSc; Cecilia Casetta[10], MD; Alexandra Papamichail[3], PhD; Brynmor Lloyd-Evans[1], PhD; Alan Simpson[3], PhD; Nick Sevdalis[2,3], PhD; Fiona Gaughran[10,11*], MD; Sonia Johnson[1,6*], DM

[1]NIHR Mental Health Policy Research Unit, Division of Psychiatry, University College London, London, United Kingdom

[2]Centre for Implementation Science, King's College London, London, United Kingdom

[3]NIHR Mental Health Policy Research Unit, King's College London, London, United Kingdom

[4]Centre for Health Services Research, City, University of London, London, United Kingdom

[5]King's Improvement Science, King's College London, London, United Kingdom

[6]Camden and Islington NHS Foundation Trust, London, United Kingdom

[7]Kingston iCope, Camden & Islington NHS Trust, London, United Kingdom

[8]Centre for Outcomes Research and Effectiveness, Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

[9]Division of Psychiatry, University College London, London, United Kingdom

[10]Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, United Kingdom

[11]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Rebecca Appleton, PhD
NIHR Mental Health Policy Research Unit
Division of Psychiatry
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 2039872554
Email: r.appleton@ucl.ac.uk

Abstract

Background: Early in 2020, mental health services had to rapidly shift from face-to-face models of care to delivering the majority of treatments remotely (by video or phone call or occasionally messaging) due to the COVID-19 pandemic. This resulted in several challenges for staff and patients, but also in benefits such as convenience or increased access for people with impaired mobility or in rural areas. There is a need to understand the extent and impacts of telemental health implementation, and barriers and facilitators to its effective and acceptable use. This is relevant both to future emergency adoption of telemental health and to debates on its future use in routine mental health care.

Objective: To investigate the adoption and impacts of telemental health approaches during the COVID-19 pandemic, and facilitators and barriers to optimal implementation.

Methods: Four databases (PubMed, PsycINFO, CINAHL, and Web of Science) were searched for primary research relating to remote working, mental health care, and the COVID-19 pandemic. Preprint servers were also searched. Results of studies were synthesized using framework synthesis.

Results: A total of 77 papers met our inclusion criteria. In most studies, the majority of contacts could be transferred to a remote form during the pandemic, and good acceptability to service users and clinicians tended to be reported, at least where the alternative

to remote contacts was interrupting care. However, a range of impediments to dealing optimal care by this means were also identified.

Conclusions: Implementation of telemental health allowed some continuing support to the majority of service users during the COVID-19 pandemic and has value in an emergency situation. However, not all service users can be reached by this means, and better evidence is now needed on long-term impacts on therapeutic relationships and quality of care, and on impacts on groups at risk of digital exclusion and how to mitigate these.

Trial Registration: PROSPERO International prospective register of systematic reviews CRD42021211025; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021211025

(J Med Internet Res 2021;23(12):e31746) doi:10.2196/31746

KEYWORDS

telemental health; COVID-19; remote care; telemedicine; mental health; systematic review, implementation science

Introduction

Since the onset of the COVID-19 pandemic in the first few months of 2020, most countries have experienced a severe disruption of mental health service delivery in its usual forms [1]. Community-based outpatient, day and home treatment programs, prevention and mental health promotion programs, and services for specific age groups, such as older adults, children, and adolescents and people with substance misuse problems, have been among those severely affected at a time of potentially increased demand due to the adverse mental health consequences of the pandemic [2,3].

Mental health care providers around the world responded to the disruption of services in many ways, including the significant and widely documented shift to remote delivery of mental health services to replace in-person consultations [1,4,5]. Telemental health, defined as "the provision of behavioral and/or mental health care services using technological modalities in lieu of, or in addition to, traditional face-to-face methods" [6], including video conferencing, telephone, email or text messaging, has been central to continuing assessment and support in the community. Additionally, technological innovations are helping to address isolation and service disruption in hospital and residential settings [4,7].

Multiple research studies conducted both before and during the pandemic have reported evidence of the effectiveness of telemental health in reducing treatment gaps and improving access to care for a range of service users [8-10]. Findings from studies, often of telemental health programs established for purposes of research, have suggested that, overall, synchronous modalities such as video conferencing are comparable to face-to-face delivery in terms of quality of care, reliability of clinical assessments, and treatment outcomes and adherence [11-15]. Good levels of service user acceptance and satisfaction with telemental health services have also been reported [10]. Successful adoption of telemental health has been described across a wide range of populations (adult, child and adolescent, older people, ethnic minorities), settings (hospital, primary care, community), and conditions [11,13,16]. For certain populations, including some with autism and severe anxiety disorders, and those with physical disabilities or geographical barriers to accessing services, telemental health can be preferable for some service users [6,17], although individuals experiencing significant social disadvantage or severe mental health problems, such as psychosis, have been found to benefit less [18]. Research suggests that telemental health can also work for group interventions [19]. The attitudes of clinicians who have delivered care via synchronous telemental health appear to be largely positive, with professionals finding it both effective and acceptable [20] and recognizing its potential to enhance communication within and between mental health teams [4,7]. There is also some positive health economic evidence, with several studies suggesting telemental health is no more expensive than face-to-face delivery and tends to be more cost-effective [12]. This approach also appears to be a viable and inexpensive treatment option where access to emergency services is limited, and associations have been found with reduced psychiatric admissions [10].

However, despite this evidence base, integration of telemental health approaches into routine mental health care or the widespread adoption of remote working across whole systems has rarely been reported. Even during the pandemic, adoption of such technologies has been piecemeal, with utilization varying substantially both between and within countries [1,7]. Technological barriers to the wider adoption of telemental health include (1) the risk of digital exclusion of some service users, such as those facing significant social disadvantage or with limited technological access and expertise, and (2) the lack of technological infrastructure and clear protocols within services, impeding the integration of telemental health with face-to-face care [4,21,22]. Other barriers include difficulty in establishing and maintaining therapeutic relationships and in conducting high-quality assessments; service users who lack private space or find participating in sometimes intimate and distressing discussions from home intrusive [4,11,12,18,21-23]. A range of other ethical, regulatory, technological, cultural, and organizational barriers have also been identified, both before and during the pandemic [12,24-27].

The widespread emergency adoption of telemental health since the onset of the pandemic has generated a substantial literature. Numerous commentaries, service evaluations and reports of telemental health innovations, and service user, carer and staff experiences, in addition to a growing number of research studies addressing effectiveness and implementation issues [28-31] have been published internationally. Clinical guidelines have been rapidly produced in a number of countries [32].

A synthesis of the relevant empirical evidence gathered during the pandemic is therefore timely and informative for planning by generating evidence of effects of adopting telemental health across whole populations and service systems rather than in the context of relatively small-scale research studies involving volunteer participants. Capturing the learning and experiences gained through the rapid shift to telemental health will help optimize remote health care in a population that presents unique relational challenges associated with mental distress. It will also help to understand and overcome implementation barriers and inform strategies for improving the flexibility, effectiveness, and efficiency of mental health services through the sustained integration into routine care of telemental health approaches, to ensure that it brings the greatest benefits for patients, carers, and staff.

The aim of this review is to synthesize the international literature specific to remote working in mental health services (as a replacement for or in conjunction with face-to-face service delivery) in the context of early stages of the COVID-19 pandemic. The paper complements our previous umbrella review (systematic review of reviews) of literature on telemental health prior to the COVID-19 pandemic [15]: focusing on the literature from the pandemic period allows us to identify specified learning from the very wide implementation that occurred during this period. Our research questions are as follows:

1. What evidence has been obtained during the COVID-19 pandemic regarding the effectiveness and cost-effectiveness of telemental health and regarding its safety (including adverse events due to breaches of privacy and safety)?
2. What coverage has been achieved through telemental health adoption in the pandemic (including extent of adoption by services and reach among clinical populations in which it is adopted); in which groups and for which service settings is telemental health more or less likely to be implemented successfully; what are potential risks associated with digital exclusion for those not reached; and what barriers and facilitators influence success in implementation?
3. How acceptable are telemental health approaches to service users, carers, and staff as applied during the pandemic, including perceived impacts on therapeutic relationships, communication, and privacy?
4. What innovations and improvements have been introduced to make clinical care via telemental health more effective and acceptable, achieve greater coverage, and address barriers to delivering care in this way? (This includes descriptions and evaluations of specific strategies designed to make telemental health work better than usual delivery, and of adaptations of telemental health to specific settings, such as inpatient wards and crisis services).

Methods

A systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [33]. The protocol for this review was registered with PROSPERO (CRD42021211025).

Search Strategy

Four electronic databases (PubMed, PsycINFO, CINAHL, and Web of Science), preprint servers medRxiv, PsyArXiv, Wellcome Open Research, and JMIR were searched for research relating to COVID-19, mental illness, and remote working from January 1 to December 9, 2020. An example search strategy can be found in Multimedia Appendix 1.

This search was supplemented by searching the references listed in included studies for any additional studies that met our inclusion criteria.

Screening

The resulting list of articles was deduplicated using Endnote [34] and all references were imported into Rayyan [35] for title and abstract screening. Full texts were sourced for articles deemed relevant for inclusion, and these were screened against the full review eligibility criteria. To establish consistency in study selection, title and abstract screening was conducted by 4 reviewers (MS, ZH, JH-S, and LSR), with 100% of included and 25% of excluded references checked by another reviewer (RA). Full texts were screened by 3 reviewers (RA, MB, and MS), with 100% of included and 25% of excluded papers checked by another reviewer (task divided between LG, HJ, JW, PB, and LSR). Any disagreements were resolved through team discussion.

Inclusion Criteria

Participants

Staff working within the field of mental health, people receiving organized mental health care for any condition (including addictions, dementia, and intellectual disability), family members or carers of people receiving mental health care (regarding their views on the impact of remote working on the service user, and interventions aimed at reducing carer distress). There are no age restrictions on participants in this review.

Interventions

Any form of spoken or written communication carried out between mental health professionals or between mental health professionals and service users/family members/unpaid carers or peer support communications using the internet, the telephone, text messaging platforms, or hybrid approaches combining different platforms.

Comparator(s)/Control

Any mental health communication delivered face-to-face, digitally or remotely, waitlist control, or placebo. Studies comparing different modes of delivery during the pandemic, and those comparing care delivery and outcomes during the pandemic with those before the pandemic were included. Relevant studies with no comparator were also included.

Outcomes

Qualitative and quantitative outcomes describing implementation effectiveness (including process evaluations) and barriers and facilitators to digital engagement, clinical effectiveness, cost-effectiveness, acceptability (including service user, carer, and staff satisfaction), impacts on communication and

therapeutic relationships, coverage and impacts of digital exclusion, interventions to improve quality or coverage, improvements in quality of life, and economic impacts.

Design

Any papers that present qualitative or quantitative data from study designs of any type (including relevant service evaluations and case series). If the focus of the study was not solely remote working but the results section contained substantial data relevant to our research questions, these were also included. Any relevant reviews identified in the searches were checked for included research which met our inclusion criteria.

Exclusion Criteria

We excluded studies that were (1) not specific to the pandemic response; (2) reporting on interventions with patients with primary sleep disorders; (3) reporting on those with subclinical symptoms (unless combined with another included mental health problem); (4) focused on digital interventions such as apps, websites, and virtual reality tools, except where the sole purpose of the digital intervention was to facilitate direct spoken or written communication; (5) focused on interventions aimed at improving the mental health or well-being of health care professionals; and (6) editorials, opinion pieces, guidance documents, protocols, conference abstracts, and letters, with the exception of editorials or letters which contained primary research findings.

No language or location restrictions were applied in this review.

Data Extraction

Data extraction was supported by well-established implementation science frameworks. A data extraction form was developed based on a brief version of the Consolidated Framework for Implementation Research (CFIR) [36] and the taxonomy of implementation outcomes [37]. We used the higher-level CFIR constructs (see Table 2 in section *Barriers and Facilitators to Telemental Health* for a brief definition of each one of the implementation facets that CFIR constructs capture) to extract data on factors influencing implementation success [38], and the taxonomy of implementation outcomes including acceptability, adoption, and feasibility. We also extracted information deemed relevant based on previous studies conducted by the research team, including an umbrella review of pre-COVID-19 systematic reviews on telemental health and a qualitative study [15,21]. Data extracted consisted of study details, including design and focus of study; gender, ethnicity, age; diagnosis of participants; details of staff occupation; setting and context of study; intervention details, implementation outcomes (including acceptability, adoption, appropriateness, feasibility, fidelity, cost effectiveness, penetration, and sustainability); barriers and facilitators to implementation; and clinical and safety outcomes. The full data extraction form can be viewed in Multimedia Appendix 1.

Data extraction was completed by 9 reviewers (AP, JW, MS, ER, RA, MT, JM, SS, HJ, approximately 8 studies each) using EPPI-Reviewer 4 [39]. All reviewers were trained on how to extract data to ensure consistency, and extracted data were checked by a second reviewer (RA & NVSJ). The extraction form was first piloted on 9/77 (12%) of included studies to assess usability and content, with amendments made before completing extraction for the whole data set.

Quality Appraisal

Given the diversity of the included article types and methods, 2 quality appraisal tools were used. Primary research studies were assessed with the Mixed Methods Appraisal Tool (MMAT) [40]. Commentaries and service evaluations were assessed using AACODS (authority, accuracy, coverage, objectivity, date, significance) tool, which appraises the veracity, clarity, acknowledgement of bias, and relevance of the contribution to the field [41]. Study quality was assessed by RA and verified by NVSJ. Disagreements were resolved through discussion.

Evidence Synthesis

We conducted a framework synthesis of study characteristics and outputs. Study outcomes were tabulated by applying existing implementation science frameworks, that is, the CFIR framework [36], Proctor et al's [37] taxonomy of implementation outcomes; and also by relevant topics/themes that emerged during data extraction. This table-based synthesis of the study outcomes combined a deductive and inductive approach to data analysis by using existing frameworks, while identifying emerging themes. Results reported in this paper include a narrative synthesis of the study characteristics and quantitative study outputs [42] and the tabulated results.

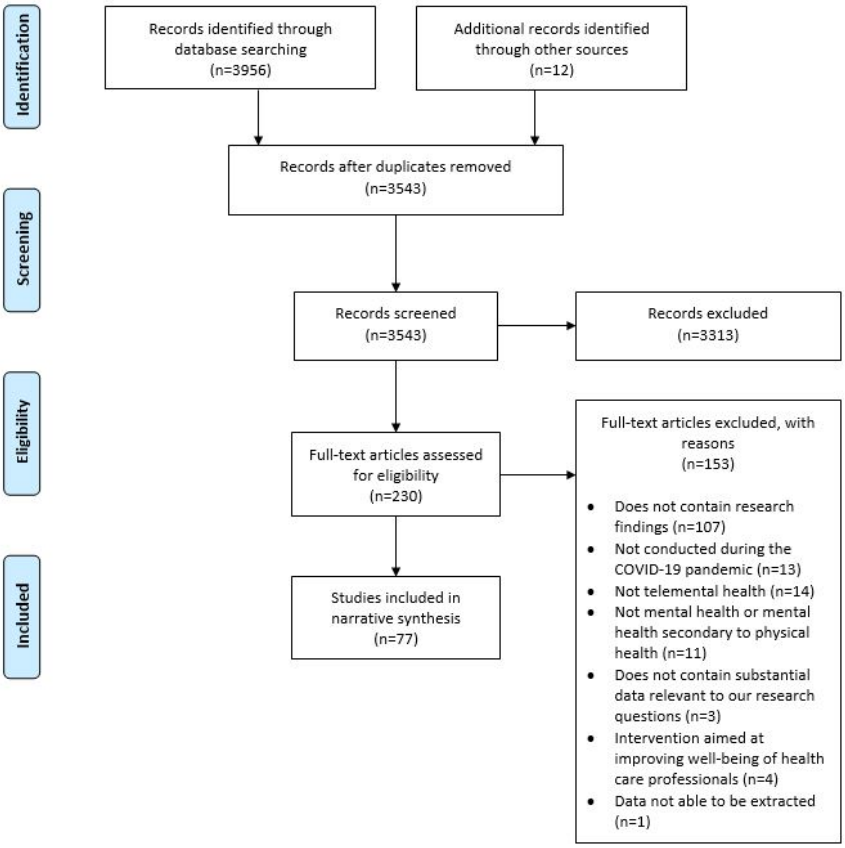
Results

Study Selection

A total of 3956 references were identified through searching databases of published papers. medRxiv was the only preprint database where included papers were found (n=10); 1 more relevant paper was identified by a member of the research team and a further paper was found through reference searching of included studies.

A PRISMA flow diagram [33] of the screening and selection process is presented in Figure 1.

Figure 1. PRISMA diagram showing screening and included studies.



Quality of Included Studies

The quality of included primary research studies was moderate to high: 23 out of 48 studies appraised using the MMAT met above 80% of quality criteria, whereas 21 out of 48 met between 50% and 79%. The quality of included service evaluations or audits was generally high: 27 out of 29 studies appraised using the AACODS met at least 4 out of the 6 quality criteria assessed. These include being written by recognized experts; including reference lists; having a clear aim; stating details such as date, location, and limitations; and making a meaningful contribution to the research literature.

Study Characteristics

Of the 77 studies which were eligible for inclusion in the review (Table 1), 45 were primary research studies, 24 were service evaluations or audits, and 8 were editorials or letters that included data. Thirty-three were conducted in the United States, 9 in the United Kingdom, and 5 each in Australia, Canada, and Spain. Five were conducted across more than 1 country.

Of the 45 primary research studies, 32 involved staff and 9 involved service users. The remaining 4 analyzed service use data (3 evaluated contacts with hotlines and 1 evaluated service use in 1 UK National Health Service [NHS] service provider).

Most studies were conducted in services that worked with people with mixed psychiatric diagnoses (n=30), although we also found studies conducted with groups with a single diagnosis (eg, dementia or eating disorders). Studies could include more than 1 service type, the most commonly studied being community mental health teams (CMHTs) or outpatient settings (n=39), followed by psychology or psychotherapy services (n=17). Inpatient or residential services were included in 15 studies, while 12 included general hospitals. Eight studies included private hospitals or clinics, while 4 explored telemental health use in helplines, voluntary sector services, crisis teams, or veterans' health services, respectively. Five studies did not report any specific setting.

The aims of most studies were either a description of changes made due to the pandemic, new services set up because of the pandemic, or an evaluation of the impact of the pandemic on either staff or service users. The descriptions of changes either focused specifically on the move to telemental health or were wider descriptions of changes to services including the use of telemental health. The characteristics of each of the included studies are shown in Table 1, with a more detailed summary in Multimedia Appendix 2.

Table 1. Study characteristics.

Study	Aim of study	Modality used	Mental health problem/diagnosis	Participants (sample size)
Aafjes-van Doorn et al [43]	Survey of therapists' experiences of video therapy during the pandemic	V[a]	Not stated	Staff (n=144)
Anton et al [44]	Description of transition to telemedicine	V, P[b], TM[c], E[d], M[e]	Depression, posttraumatic stress disorder	Staff
Barney et al [45]	Description of transition to telemedicine	V	Mixed	Staff
Békés et al [46]	Survey of psychotherapists' attitudes toward online psychotherapy	V	Mixed	Staff (n=145)
Békés et al [47]	Survey of psychoanalytical therapists' experiences of videoconference therapy during the pandemic	V, P	Mixed	Staff (n=190)
Benaque et al [48]	Description of service changes due to the pandemic	V, P, TM	Dementia	Staff
Berdullas Saunders et al [49]	Description of the use of a psychological helpline	P	Mixed	General population (15,170 calls)
Bhome et al [50]	Survey of staff perspectives on delivery of services to older adults during the pandemic	NA[f]	Dementia	Staff (n=158)
Bierbooms et al [51]	Interviews with health professionals on the sustainability of online treatment after the pandemic	V, TM, O[g]	Mixed	Staff (n=11)
Boldrini et al [52]	Survey of psychotherapists' experience with telepsychotherapy during the pandemic	V, P, M	Mixed	Staff (n=308)
Burton et al [53]	Interviews with people with mental health conditions on their experience during the pandemic	NA	Mixed	Service users (n=22)
Carpiniello et al [54]	Survey to explore the impact of the pandemic on the functioning of mental health services	NA	Mixed	Staff (n=71)
Cheli et al [55]	Evaluation of a crisis intervention for patients diagnosed with psychosis	V	Psychosis and bipolar	Service users (n=6)
Chen et al [56]	Description of changes made to mental health services due to the pandemic	V, P	Mixed	NA (description of service change)
Childs et al [57]	Description of changes made in an outpatient psychiatric service due to the pandemic	V, P	Mixed	Service users
Colle et al [58]	Evaluation of teleconsultation during the pandemic	V, P	Mixed	Service users
Connolly et al [29]	Description of changes to services during the pandemic	V, P	Mixed	NA (description of service change)
Datta et al [59]	Description of transition to telehealth during the pandemic	V	Eating disorders	NA (description of service change)
Dores et al [60]	Exploration of mental health professionals' attitudes regarding information and communications technology use	V, P, O, E	Mixed	Staff (n=108)
Erekson et al [61]	Exploration of use of telehealth in a student counseling service during the pandemic	V	Mixed	Staff
Feijt et al [62]	Exploration of staff experiences of online treatment during the pandemic	V, P, O, E	Mixed	Staff (n=51)

Study	Aim of study	Modality used	Mental health problem/diagnosis	Participants (sample size)
Fernandez et al [63]	Survey on the impact of the pandemic for people diagnosed with an eating disorder	NA	Eating disorders	Service users (n=121)
Foye et al [64]	Exploration of the impact of the pandemic on mental health nurses	V, P	Mixed	Staff
Gaddy et al [65]	Exploration of the impact of the pandemic on music therapy professionals	V, P	NA	Staff (n=1196)
Gillard et al [66]	Exploration of the experiences of people with mental health problems during the COVID-19 pandemic	V, P, TM	Mixed	Service users
Gomet et al [67]	Description and review of the implementation of remote working in an addiction outpatient service	P	Substance abuse	Service users
Graell et al [68]	Exploration of the impact of the pandemic on a child and adolescent eating disorders service	V, P, M	Child and adolescent eating disorders	Service users (n=365)
Grover et al [69]	Evaluation of the monitoring of patients with schizophrenia on clozapine during the pandemic	P, TM	Psychosis and bipolar	Service users
Grover et al [70]	Evaluation of the impact of the pandemic on mental health services in India	V, P	Mixed	Staff (n=396)
Grover et al [71]	Evaluation the impact of the pandemic on mental health services in India	V, P	Mixed	Staff (n=109)
Haxhihamza et al [72]	Evaluation of the satisfaction of patients with telepsychiatry due to the pandemic	M	Mixed	Service users (n=28)
He et al [73]	Evaluation of a psychological intervention program	V, P, M, O	General population	NA
Hom et al [74]	Description of the development of a virtual program for an acute psychiatric population	V	Mixed	Staff and service users
Humer et al [75]	Survey of psychotherapists' views on working during the pandemic	V, P, E, M	NA	Staff (n=338)
Humer et al [76]	Survey of psychotherapists view on the use of the internet during the pandemic	V	NA	Staff (n=1547)
Izakova et al [77]	Survey of the impact of the pandemic on mental health experts	V, P	NA	Staff (n=157)
Johnson et al [7]	Survey of the experiences of mental health staff during the pandemic	V, P, M	Mixed	Staff (n=2180)
Jurcik et al [78]	Exploration of how the pandemic affected mental health services	V, P, M	Mixed	Staff (n=8)
Khanna et al [79]	Description of services changes in a trauma service during the pandemic	V, P, M	Posttraumatic stress disorder	Staff (n=21)
Kopec et al [80]	Description of the transition to telehealth in a community mental health service	V, P	Mixed	NA (description of service change)
Lai et al [81]	Evaluation of the benefits of telehealth to people with dementia and their carers	V, P	Dementia	Service users (n=60)
Lakeman and Crighton [82]	Exploration of providing dialectical behavior therapy using telehealth technology	V, P, M	Personality disorder	Staff (n=28)
Lin et al [83]	Evaluation of psychological hotline services set up during the pandemic	P, TM	General population	NA (evaluation of a new service)

Study	Aim of study	Modality used	Mental health problem/diagnosis	Participants (sample size)
Looi et al [84]	Evaluation of the use of psychiatry telehealth in smaller states	V, P, M	Mixed	Staff
Looi et al [85]	Evaluation of the use of psychiatry telehealth in larger states	V, P, M	Mixed	Staff
Lynch et al [86]	Description of change to telehealth in a service for people with psychosis	V	Psychosis and bipolar	Service users (n=64)
McBeath et al [87]	Exploration of the experiences of psychotherapists working remotely during the pandemic	V, P, TM, E	Mixed	Staff (n=335)
Medalia et al [88]	Description of the change to telehealth in a service for people with serious mental illness	V	Mixed	NA (description of service change)
Miu et al [89]	Evaluation of the engagement with telehealth of people with severe mental illness during the pandemic	V, P	Mixed	Staff (n=24)
Olwill et al [90]	Survey of psychiatrists' experience of remote consultations	P	Mixed	Staff (n=26)
Patel et al [91]	Analysis of health record data on the impact of remote consultation during the pandemic	NA	Mixed	NA (description of whole service)
Peralta et al [92]	Evaluation of the effectiveness of teleconsultation use during the pandemic	V, P, TM	General population	NA (6800 interventions)
Pierce et al [93]	Survey of the impact of telepsychology use by psychologists before and during the pandemic	V, P	Mixed	Staff (n=2619)
Probst et al [94]	Investigation of changes to psychotherapy compared with the months before the pandemic	P	Mixed	Staff (n=1547)
Reilly et al [95]	Survey to understand change in practice by health care staff during the pandemic	NA	Mixed	Staff (n=903)
Roach et al [96]	Interviews to understand the experience of people with dementia during the pandemic	NA	Dementia	Service users (n=21)
Roncero et al [97]	Description of the response of a mental health network to the pandemic	V, P, M	Mixed	NA (description of service change)
Rosen et al [98]	Description of transition to telemental health services	V, P	Mixed	NA (description of service change)
Sasangohar et al [99]	Description of implementation of telepsychiatry in a psychiatric practice	V, P, E	Mixed	NA (description of service change)
Scharff et al [100]	Description of changes made by a psychological service during the pandemic	V	Mixed	NA
Schlegl et al [101]	Survey to investigate the impact of the pandemic on patients with bulimia nervosa	NA	Eating disorders	Service users (n=55)
Sciarrino et al [102]	Description of providing trauma-focused treatment using telehealth during the pandemic	V, O	Posttraumatic stress disorder	NA
Sequeira et al [103]	Description of change to services for people with obsessive compulsive disorder during the pandemic	V	OCD	Service users (n=5)
Severe et al [104]	Survey of patients using a mental health service to explore decisions to accept or decline telepsychiatry	V, P	Mixed	Service users (n=244)

Study	Aim of study	Modality used	Mental health problem/diagnosis	Participants (sample size)
Sharma et al [28]	Description of the implementation of a home-based telemental health service during the pandemic	V, P	Child and adolescent services	Staff (n=105)
Sheehan et al [105]	Survey of the experiences of staff working with people with intellectual and other developmental disabilities	NA	Intellectual disabilities	Staff (n=648)
Sklar et al [106]	Exploring the impact of the pandemic on mental health services in Indiana	NA	NA	Staff
Termorshuizen et al [107]	Survey to evaluate the impact of the pandemic on people with eating disorders	NA	Eating disorders	Service users (n=1021)
Uscher-Pines et al [108]	Interviews with psychiatrists to understand how change in delivery has affected mental health care	V, P	Mixed	Staff (n=20)
Uscher-Pines et al [109]	Interviews with clinicians to understand the experience of using telemedicine for opiate use disorder	V, P	Opiate use disorder	Staff (n=18)
van Dijk et al [110]	Description of transforming a day-treatment program for older people into an online program	V	Mixed	Staff
Wang et al [111]	Survey to compare Chinese and US practitioners' attitudes toward teletherapy during the pandemic	V, P	Mixed	Staff (n=329)
Wilson et al [112]	Survey to explore staff perceptions of the impact of the pandemic on perinatal services	V, P, M	Perinatal services	Staff (n=363)
Wood et al [113]	Description of the implementation of group teletherapy for people with first-episode psychosis	V	Psychosis and bipolar	Service users (n=7)
Wyler et al [114]	Exploration of the experience of therapy sessions for people with attention deficit hyperactivity disorder and their therapists during the pandemic	V, P	Attention deficit hyperactivity disorder	Staff and service users (n=60 therapist/service user dyads)
Yellowlees et al [115]	Description of the rapid conversion of an outpatient psychiatric clinic to a telepsychiatry clinic	V, P	Mixed	NA (description of service change)
Zulfic et al [116]	Audit to understand the move to telephone support for people using a community mental health team	P	Mixed	Service users (n=314)

[a]V: video.

[b]P: phone.

[c]TM: text message.

[d]E: email.

[e]M: mobile.

[f]NA: not applicable/not stated.

[g]O: other.

Data Synthesis

Barriers and Facilitators to Telemental Health

Implementation barriers and facilitators were categorized using a condensed version of the CFIR framework (see Table 2, where definitions of the CFIR constructs are also provided). The key findings are summarized below.

Table 2. Implementation barriers and drivers for telemental health grouped according to condensed CFIR[a] domains.

CFIR domain	Findings	Example references
Intervention characteristics: Whether the intervention was internally/externally developed, evidence supporting the intervention, advantages compared with other methods of delivery, adaptability, trialability, and complexity	• Remote care had advantages over face-to-face, for example, making therapy more accessible for certain groups such as service users in remote locations; saving users money on travel; helping therapists get a better idea about the service users' home environment; some users benefitted from the distance, found it easier to communicate openly, and became more independent. • The main barriers for clinicians to deliver quality therapy were picking up on nonverbal cues, assessing mental health symptoms, keeping service users engaged. • Video and phone calls were the most common modalities; however, studies also reported the use of emails, instant messaging services, apps, videos, and forums. • Duration of telemental health appointments were shorter than face-to-face; clinicians reported it required more concentration and was more tiring. • In some cases, studies have reported using shorter but more frequent appointments to deal with challenges in remote working (eg, some service users struggling to stay focused). This was also used as a method to increase flexibility. • Frequent contacts between sessions helped to build the therapeutic relationship.	[43,51,56,86,108,109]
Outer setting: Information on whether the organization is networked with others, peer pressure to implement intervention, and external policies and incentives	• Implementation was commonly due to "stay at home" orders or national lockdowns, or a high level of COVID-19 cases in that area resulting in social distancing requirements. • In the United States, health insurers did not always cover telemental health care, whereas in some European countries, insurance cover for telemental health terminated at the end of the first wave of infections. • Telehealth service delivery was eased by the relaxation of policy and billing reimbursements during this time. • Professional bodies facilitated transition to telehealth by posting guidelines on their websites to assist clinicians. • Platform developers worked rapidly to increase capacity. • Clinicians identified the need for a video tool that adheres to privacy standards and links with a technical helpdesk. • There were also concerns over the reduction in services to support the physical health needs of mental health service users.	[29,56,57,75]
Inner setting: Information on the structural characteristics, networks and culture of an organization, as well as the implementation climate (eg, capacity for change)	• Overall, all settings had sufficient capacity to shift to some delivery of telemental health in a short period.	[76,100,102,105]
Staff characteristics: Information on the following psychological attributes and also on any effects of staff demographic and professional backgrounds	• There was some variation in acceptability of remote ways of working for staff depending on their therapeutic approaches. • Telemental health take-up was dependent on perceived experience of patient (positive or negative), comfort with online platform, previous clinical experience. • Some staff felt less confident about professional skills during online compared with in-person consultations, especially those with less clinical experience and those who perceived their patients disliked remote care.	[43,52,111]
Process: Training provided and any processes put in place to support telemental health intervention, planning, and feedback on progress of implementation	• The transition to telemental health occurred usually over a short period. • Training staff to use platforms was mentioned frequently, as was phoning service users to let them know about the transition to telemental health and how care would be provided going forward. • Methods of staff training included courses, shadowing or observing senior colleagues, discussion within clinical teams, facility-level telehealth coordinators, clinical champions providing training, and webinars. • Sources of information for staff: colleagues, government guidelines, prepared consent forms, posts on listservs, American Psychological Association, and other official guidelines. • New workflows had to be developed to allow staff to access patient records remotely. • Despite some training, staff reported lack of support and identified training needs across several studies regarding how to use online platforms and meeting privacy regulations.	[7,28,44,45,50,59,61,62,74,105,112,115]

CFIR domain	Findings	Example references
Service user needs/resources: Statements demonstrating awareness of the needs and resources of those served by the organization (eg, barriers and facilitators and feedback)	• A commonly reported issue was access to technology, particularly among service users with diagnoses such as schizophrenia, service users with a lower socioeconomic status, and older adults (one study mentioned that older adults often lacked access to video software, so preferred phone calls). • Concerns around privacy and confidentiality, and forming a therapeutic relationship may be more difficult when using remote care. • Difficulties for service users to concentrate within a digital environment. • Several studies mention the need for an agreed "Zoom etiquette" for service users, including attire, audio/visual setup, and reducing background distractions. • Stable internet connection was a problem for some service users. • Some clients benefitted from the distance created by online treatment, as they became less inhibited and less dependent on therapist.	[50,62,77,86,98,116]

[a]CFIR: Consolidated Framework for Implementation Research.

Intervention Characteristics

Video and phone calls were the most common modalities used for remote care; studies also reported the use of emails, instant messaging services, apps, prerecorded videos and forums (further details about the modality used in each study can be found in Table 1).

When comparing remote care with traditional face-to-face settings, studies identified advantages for both methods. Benefits for remote care included being more convenient (for both staff and service users), making care more accessible to groups who may previously have been excluded, reducing travel (resulting in both time and cost savings), and helping clinicians understand more about the service user, as they had more insight into their home lives. A further benefit is that more family members were readily able to attend family therapy or family education sessions since care was moved online (eg, [43]). However, clinicians reported difficulties in picking up on nonverbal cues in remote compared with face-to-face care, and that remote care could sometimes require more concentration.

Outer Setting

Services commonly implemented remote methods of working due to "stay at home" orders or national lockdowns, or due to a high level of cases in their local area. Overall, all settings described in papers had sufficient capacity to make a rapid shift to remote forms of care. Several studies in the United States in particular mentioned the impact of health insurance regarding uptake of telemental health (eg, [56]), as not all insurance providers covered remote care. However, this did change during the course of the pandemic as telemental health delivery was eased by the relaxation of policy and billing reimbursements [56,75]. The change from face-to-face to remote delivery of care was also facilitated by professional societies who posted guidelines on their websites to assist clinicians.

Staff Characteristics

Enablers for clinician uptake included supporting clinicians by ensuring supervision, supportive leadership, clear communication, keeping track of clinicians' needs, optimizing physical space for comfort and privacy (eg, using headphones or ergonomic seating), and arranging times away from the computer.

However, staff in several studies reported a lack of initial training for telemental health, and therefore identified training needs regarding the use of online platforms and meeting privacy regulations in particular. In some studies, having no previous experience with telemental health was also found to be associated with higher anxiety [43] and lower uptake [52] of remote care. However, others found that previous experience did not impact clinicians' views of telemental health during the pandemic [47].

Process

As telemental health was not commonly used in most services before COVID-19, staff had to rapidly adjust to a new way of working. Several studies discussed the training which was put in place for staff, which included training courses, shadowing, or observing senior colleagues; discussion within clinical teams' facility-level telehealth coordinators and clinical champions providing training; webinars; and checking official guidelines. New workflows also had to be developed to allow staff to access patient records remotely, and service users had to be informed about the transition to telemental health.

Service User Needs/Resources

In addition to the needs of staff, service users also identified certain needs and resources to enable them to effectively transition to telemental health care. A commonly reported issue was access to technology, particularly among service users with diagnoses such as schizophrenia (eg, [116]), older adults (eg, [50]), and service users from lower socioeconomic backgrounds (eg, [78]). Service users also reported problems having a stable internet connection to allow for uninterrupted communication, which could negatively impact the therapeutic relationship. Concerns were also raised by both clinicians and service users regarding privacy and confidentiality, and in some cases service users had difficulties concentrating on remote care. Several studies (eg, [86,98]) mentioned the need for an agreed "Zoom etiquette" for service users, including attire, audio/visual set up, and background distractions.

Implementation Outcomes

Outcomes of the implementation of telemental health have been summarized below using Proctor et al's [37] taxonomy of implementation outcomes. Further information can be found in Table 3.

Table 3. Implementation outcomes summary findings for telemental health.

Implementation outcome	Findings	Example studies
Acceptability	• Remote methods of care are acceptable to most service users and "exceeded expectations" in terms of satisfaction, but are not viewed as a substitute for face-to-face care. • Clinicians and service users consider the intimacy and connection of face-to-face care are not reproducible on virtual platforms, especially for treatments involving nonverbal communication. • Beyond the pandemic: further data are needed about longer-term acceptability, observance, quality of care, and satisfaction. • Clinician burnout due to more appointments per day and requiring more concentration.	[7,45,60,64,70,74,76,78,87,91,94,96,97,100,104]
Adoption	• Remote working was generally well adopted (most service users switched to remote working). • A few studies also mentioned lower levels of cancellations/no shows, likely due to not having to travel to the service and the removal of other barriers (eg, difficulty fitting care around school or work). • Remote working also had the potential to result in reduced waiting times. • Productivity was generally maintained, or in some cases even increased. • Some studies showed no decrease, just change in modality and need to modify psychological treatment.	[7,50,52,54,56,58,61,75,79,81,82,113,114]
Appropriateness	• Difficulties managing medication prescription during online consultations. • Concerns around user engagement and assessing new patients. • Harder to assess mental status markers such as hygiene or eye contact, or physical symptoms (eg, of opioid withdrawal). Although it allows to know more about home environment and behavior outside of clinic. • Does not capture the richness of in-person interaction. • Online felt safer for clinicians providing care to service users at risk for violence and behavioral dysregulation. • Not appropriate for patients with auditory or visual impairments, or with conditions such as migraines.	[56,60,78,82,99,100,112]
Feasibility	• Links with service user and staff needs and resources, in particular problems accessing technology/private space/stable internet connection. • All studies reported good feasibility at least for the short-term emergency response during the pandemic. • However, it was not possible to use for specific therapies that require physical presence (role play, collaborative models). Telemental health was less suitable for treating trauma, for clients with severe anxiety, children, and clients with cognitive impairment. • Insurance coverage and legal aspects affected feasibility of implementation in some countries. However, most health insurances caught up and started covering costs.	[7,56,58,62,64,73,74,78,82,94,96,104]
Fidelity	• No studies explored this area.	
Implementation cost	• Limited information about cost of intervention, suggested to be "cost effective" without any presentation of costs. • Reduced travel costs.	[59,70,87,103]
Penetration	• Prior to the pandemic, few services used telemental health and for those that did, uptake was low. After the first few weeks, most or all of services were conducted remotely.	[7,54,105]

Implementation outcome	Findings	Example studies
Sustainability	• Rates of telemental health use fell as COVID-19 rates declined in the summer of 2020. Links with findings that not all staff and service users would want to continue using remote methods of care after the pandemic ends. • Flexibility is a key advantage of telemental health versus face-to-face care. • There are some aspects of remote working that services would like to keep, as they provide benefits such as being more efficient and enabling access for certain groups. • Some barriers to remote working (such as lack of experience with online methods of care) have been reduced, making it more likely telemental health will continue to some extent.	[43,64,85,93,95,104]

Acceptability

Remote care was seen as satisfactory by the majority of clinicians and service users in most studies in the context of the pandemic. A number of studies also reported that telemental health enabled some groups to access care who found it difficult to engage with face-to-face support (eg, [7]). Some clinicians reported that they would also be willing to continue with some aspects of remote care in the future (eg, [43,77]). However, it is important to note that while acceptability was high overall, this was not the case for all groups; for example, Grover et al [71] reported acceptability rates of around 45% for both clinicians and service users using services in a range of settings in India. Further details of satisfaction and acceptability outcomes are presented in Table 4.

Telemental health services were acceptable to people during the pandemic as a way of continuing their treatment; however, findings from several studies also indicated that participants wanted at least some appointments to be face-to-face once restrictions on in-person contact had loosened.

Table 4. Levels of acceptability of telemental health during the COVID-19 pandemic.

Author	Type of service	Service location	Acceptability data
Aafjes-van Doorn et al [43]	Psychology/psychotherapy/counseling service	United States, Canada, Europe (Hungary, Italy, United Kingdom, Germany, Norway, Sweden, Switzerland, Latvia, Ireland)	• Clinician views: Mainly positive attitudes toward video therapy were reported (mean 3.42 [SD 0.50]; range: 2.31-4.69). Views on video therapy had become more positive since the pandemic (t140=2.06, *P*<.05); video therapy was still viewed as somewhat less effective compared with in-person therapy (mean 2.19 [SD 0.65]; range: 1.00-4.00). • Service user and carer views (reported by clinicians): Only 7% (n=10) thought their patients experienced video therapy negatively. The majority perceived patient experience as either positive (N=88, 63.8%) or neutral (N=40, 28.4%).
Békés et al [46]	Psychology/psychotherapy/counseling service; private hospital/clinic; CMHT[a] and outpatient services	Canada, United States, Europe (countries not stated)	• Service user and carer views (reported by clinicians): Psychotherapists reported that their patients had an extremely positive (N=20, 13.8%), positive (N=71, 49%), or neutral (N=40, 27.6%) experience with online psychotherapy. About 7.6% of the psychotherapists thought that their patients experienced online psychotherapy somewhat negatively and none of the psychotherapists reported an extremely negative patient experience.
Békés et al [47]	CMHT and outpatient services; psychology/psychotherapy/counseling service; private hospital/clinic	Canada, United States, Europe (countries not stated)	• Clinician views: Challenges included technical/internet problems (64.7%), patients not having a private space (46.8%), risk of patient (44.7%) or therapist (26.3%) getting distracted, difficulty feeling connected to patients (29.5%) or reading their emotions (27.4%), difficulty keeping professional boundaries (23.2%), and confidentiality concerns (16.3%). About 64.2% (n=122) reported their relationships with service users felt as authentic to before COVID-19, 46% felt as emotionally connected, and 64% reported no change to the therapeutic relationship. • Service user and carer views (reported by clinicians): Most therapists reported a positive (n=101, 53.2%) or neutral (n=55, 28.9%) patient experience, with only 34 reporting a somewhat negative online therapy experience for their patients (25.8%).
Benaque et al [48]	Voluntary sector/nonprofit	Spain	• Clinician views: 81% of clinical staff considered the quality of telemedicine consultations to be either good or excellent; 75% viewed telemedicine visits as equal or better than face-to-face consultations.
Colle et al [58]	CMHT and outpatient services	France	• Clinician views: 94.1% of psychiatrists were satisfied with teleconsultations in this context. • Service user and carer views: 89.5% of patients were satisfied and 73.3% of patients spontaneously expressed their gratefulness for remote care.
Dores et al [60]	Psychology/psychotherapy/counseling service	Portugal	• Clinician views: 21 (out of 71) psychologists (29.6%) considered their experiences to be neither negative nor positive. Most of the respondents considered their experience with digital technologies to be either positive (n=37, 52.1%) or very positive (n=13; 18.3%). None reported their experiences as negative.
Grover et al [69]	CMHT and outpatient services	India	• Service user and carer views: 75.5% of patients and family members were satisfied they could remain in touch with the treating doctor. A quarter of patients had difficulty in procuring clozapine, with clozapine not being available in their locality in 15% of cases and 3.4% having to switch their brand. 25% were able to get the absolute neutrophil count done in the previous month.

Author	Type of service	Service location	Acceptability data
Grover et al [70]	CMHT and outpatient services; inpatient mental health service; private hospital/clinic	India	• Service user and carer views (as reported by clinicians): 21% reported that non-HCWs[b] in quarantine were dissatisfied, 19.9% reported that HCWs in quarantine were dissatisfied, and 13.5% reported that HCWs working with patients with COVID-19 were dissatisfied. • Clinician views: Participants rated their satisfaction with the services they were currently providing to their patients with a mean of 45.8% (SD 28.6) on a Likert scale from 0 to 100.
Grover et al [71]	Medical colleges, government-funded institutes, general hospital psychiatry units	India	• Clinician views: Overall satisfaction with the mental health services being catered; the participants rated their mean level of satisfaction as 46.6% (SD 27.6).
Haxhihamza et al [72]	Day hospital	Macedonia	• Service user and carer views: 20/28 strongly agreed/agreed that the medical care received was just about perfect; 4 patients agreed that they were dissatisfied with some things about their medical care (1 strongly agreed and 3 agreed); 20 (strongly) agreed that they can get medical care whenever they need it; 20 (strongly) agreed that they have easy access to medical specialists; 4 (strongly) agreed that the wait for emergency treatment was too long.
He et al [73]	Helplines; online media programs	China	• Service user and carer views: Feedback from clients demonstrated that more than 50% felt their negative emotions, such as anxiety and depression, were relieved.
Hom et al [74]	Private hospital/clinic	United States	• Service user and carer views: Patients who have been discharged thus far (n=10) have also expressed confidence in their aftercare plans; 2 patients who completed the exit survey reported very positive experiences and both rated their care as 9/10.
Izakova et al [77]	CMHT and outpatient services; inpatient mental health service	Slovakia	• Clinician views: 69.4% of them have considered it as an adequate form for diagnostics and therapy in the common clinical practice; 51.6% want to use it at a limited level with the defined guidelines in future.
Johnson et al [7]	All service settings, including inpatient, CMHTs, voluntary sector	United Kingdom	• Clinician views: A majority (n=818, 74.0% of respondents) agreed/strongly agreed that video calls were suitable to assess progress of existing service users, but only 39.8% (n=442) agreed/strongly agreed that they were suitable for making the initial assessments. A majority (n=725, 65.8%) agreed/strongly agreed that use of remote care had resulted in not having contact with some service users who had not engaged with remote appointments.
Lakeman and Crighton [82]	Psychology/psychotherapy/counseling service	Australia	• Clinician views: 32% (n=7) stated they were not confident at all in delivering online DBT[c], 50% (n=11) reported being "a little" confident and 4 reported feeling confident doing so; 14 respondents identified limited access to the internet, appropriate devices, or internet blackspots as being significant obstacles to engagement.
Lynch et al [86]	CMHT and outpatient services	United States	• Service user and carer views: The telehealth acceptance rates of the CP[d] subsample indicated that 90% (n=18) enrolled at the time of conversion agreed to telehealth sessions within 10 days of the service transition.
Olwill et al [90]	CMHT and outpatient services	Ireland	• Clinician views: 92% of respondents (n=24) (and 100% consultants [n=12]) reported lower confidence in making a diagnosis. 96% (n=25) agreed that the lack of visual cues affected their assessment of the patient; 70% agreed that they found it more difficult to consider discharging a patient; 88% agreed they found it more difficult to establish a therapeutic alliance with new patients.

Author	Type of service	Service location	Acceptability data
Sheehan et al [105]	CMHT and outpatient services	United Kingdom	• Clinician views: 53.3% reported concerns of having to adapt too quickly to new ways of working; 37.9% reported having to learn new technologies too quickly or without sufficient training or support; 45.3% raised concerns around engaging patients with learning difficulties or autism; 23.7% had concerns around safeguarding or risk management; 27.9% reported greater workload than usual.
Wang et al [111]	Not stated/unclear	United States and China	• Clinician views: Before COVID-19, 25% of US psychoanalytic practitioners felt mainly negative about teletherapy and 36% felt mainly positive, as compared with only about 9% and 47% of CAPA[e] practitioners, respectively; during the pandemic about 23% of US psychoanalytic practitioners felt mainly negative about teletherapy and about 37% felt mainly positive, compared with about 2% and about 58% of CAPA practitioners, respectively.
Wilson et al [112]	CMHT and outpatient services; crisis and emergency mental health services; inpatient mental health service	United Kingdom	• Clinician views: Staff reported feeling less able to assess women attending the perinatal mental health service using telemedicine, particularly their relationship with their baby (43.3%, 90/208), and to mobilize safeguarding procedures (29.4%, 62/211).

[a]CMHT: community mental health team.

[b]HCW: health care worker.

[c]DBT: dialectical behavior therapy

[d]CP: complex psychosis

[e]CAPA: China American Psychoanalytic Alliance.

Adoption

Adoption rates were relatively high across studies, with most services or clinicians moving their appointments to remote methods. Rates of adoption of telemental health for service users who were already receiving care at the start of the pandemic ranged from 48% [89] to 100% [44,67,110]. Some studies reported face-to-face appointments still took place if necessary, for example, for initial assessments or for medication reviews (eg, [116]). Most studies that examined impact on attendance reported no adverse effects on attendance rates after introducing telemental health: there was either no difference in missed appointments when comparing remote with face-to-face care [45,86], or nonattendance after adoption of telemental health decreased [79,91,103]. Further details about adoption of telemental health across studies can be found in Table 5.

While most studies reported high adoption rates, a few studies reported a decrease in attendance: for example, Erekson et al [61] (though possibly because of the university setting) and Dores et al [60] identified challenges in retention due to low client adherence, lack of privacy, interruptions at home, lack of appropriate technology, or simply preference for face-to-face contact.

There was also evidence to indicate that adoption rates of telemental health fell as COVID-19 cases decreased (eg, [84,85]).

Table 5. Levels of adoption and coverage of telemental health during the COVID-19 pandemic.

Author	Type of service	Service location	Adoption/coverage data
Anton et al [44]	General hospital/physical health service	United States	77% (n=20) of those approached via telephone enrolled in the program, higher than the observed in-person rates of 61%. 80% of patients who were contacted by phone in the hospital agreed to be enrolled, lower than the 98% success rate when staff approaches patients in person at the bedside. 100% of patients who received in person care and 100% on the waitlist (n=5) transitioned to telepsychotherapy.
Barney et al [45]	CMHT[a] and outpatient services	United States	The percentage of provider telemedicine visits increased from 0% to 97%. The number of overall clinic visits did not decline when compared with that a year before (337 visits in March 2019 vs. 332 visits in March 2020), No-show rates were comparable between remote and face-to-face care.
Békés et al [47]	CMHT and outpatient services; psychology/psychotherapy/counseling service; private hospital/clinic	Canada, United States, Europe (countries not stated)	Before COVID-19 an average of 23 sessions (SD 10.58) per week were conducted in person, 3 sessions (SD 2.28) by phone, and 1 session (SD 2.84) online via videoconferencing. During COVID-19 an average of 7 (SD 7.91) of the in-person sessions changed to sessions by phone, and 15 (SD 10.33) to online sessions.
Benaque et al [48]	Voluntary sector/nonprofit organization	Spain	Initially, average weekly visits dropped from 657 to 254 in the first week after the state of alarm was declared. This drop was of 44% for follow-up visits and 40% for on-demand consultations. By week 16 the total number of visits (n=514; 78%) was almost up to prepandemic levels.
Boldrini et al [52]	CMHT and outpatient services; psychology/psychotherapy/counseling service; private hospital/clinic	Italy	42.1% (SD 28.9) of their psychotherapy treatments were interrupted during the lockdown. The remainder of their treatments was primarily delivered via online video (63.7% [SD 38.3]) or telephone (29.1% [SD 25.3]). 7.2% (SD 15.1) of their treatments were delivered face-to-face.
Carpiniello et al [54]	CMHT and outpatient services; inpatient mental health service	Italy	75% of appointments were switched to remote. Telehealth modalities used were mainly phone calls (100% of the Italian Departments of Mental Health), videocalls (67%), or emails (19%), with 41% of units adopting all these means of contact.
Chen et al [56]	General hospital/physical health service	United States	The outpatient psychiatry division switched from under 5% virtual visits in March 2019 to over 97% in March 2020. Productivity was maintained at about 95% of previous levels, with 9206 virtual visits in March 2020. Between March 30 and April 24, 2020, 30% of the virtual visits were conducted via phone.
Childs et al [57]	CMHT and outpatient services	United States	Before COVID-19, 100% of care was delivered in person. In the first week after shutdown, telehealth comprised 65.45% of visits (100% over the telephone). In the second week, 91.6% of visits were conducted using telehealth (83.49% over the telephone and 15.6% video). By the third week 99% of appointments used telehealth (30% using video). The percentage of appointments using video increased weekly, peaking at 69.9%
Colle et al [58]	CMHT and outpatient services	France	After 2 weeks of teleconsultations, 376 (91.0%) out of the 413 previously planned appointments were performed.
Connolly et al [29]	CMHT and outpatient services; Veterans Affairs service	United States	Daily TMH-V[b] encounters rose from 1739 on March 11 to 11,406 on April 22 (556% growth, 222,349 total encounters). Between March 11 and April 22, 114,714 patients were seen via TMH-V. A total of 88,908 (77.5%) were first-time TMH-V users. A total of 12,342 mental health providers completed a TMH-V appointment between March 11 and April 22; 4281 (34.7%) were first-time TMH-V users. Daily telephone encounters rose from 6348 on March 11 to 34,396 on April 22 (442% growth). Daily in-person encounters fell from 57,296 on March 11 to 10,931 on April 22 (81% decrease).

Author	Type of service	Service location	Adoption/coverage data
Dores et al [60]	Psychology/psychotherapy/counseling service	Portugal	During the lockdown period, 17 (15.7%) of the 108 psychologists discontinued therapy and counseling; 53 (58.2%) continued to provide services to most or all of their clients; 23 psychologists (25.3%) decreased the number of clients they saw to a range of between 0% and 25%, and for another 15 psychologists (16.5%) that number diminished to a range of between 26% and 50%.
Erekson et al [61]	Psychology/psychotherapy/counseling service	United States	Attendance rates for individual therapy temporarily dropped by about 35% but climbed to previous levels within 2 weeks. Group therapy attendance dropped by about 30% but did not fully recover, remaining about 15% lower after 2 weeks. The number of clients receiving individual therapy in 2020 dropped by 43%. Between March 22 and April 4, 2020, the service had fewer than half the intakes of any other recent year.
Gaddy et al [65]	Music therapy service	United States	Of the 869 respondents indicating current contact hours, 70.54% reported that they were providing alternative services, including telehealth services (54.81%), virtual music lessons (17.01%), prerecorded songs/playlists (16.98%), and prerecorded video sessions (16.00%). Individual services increased (mean 61.58% [SD 41.26], whereas group services decreased (mean 24.97% [SD 37.56]).
Gomet et al [67]	General hospital: addiction service	France	100% of service users took part in remote care
Graell et al [68]	CMHT and outpatient services; inpatient mental health service	Spain	During the study period, a total of 1818 outpatient consultations were carried out: 1329 (73.10%) by telephone or videoconferencing and 489 (26.9%) face-to-face.
Grover et al [69]	CMHT and outpatient services	India	The majority of the patients reported that they were in touch with their treating doctor (81.5%), with contact initiated by the treating team in 79% of patients.
Grover et al [70]	CMHT and outpatient services; inpatient mental health service; private hospital/clinic	India	Use of teleservices almost doubled during the lockdown period: 206 (52%) participants provided telecommunication services during the lockdown period, 186 (47%) provided free tele-consultation to the general public, and 269 (67.9%) provided free tele-consultation to their patients; 132 (33.3%) were using both voice and video calls (combination of free and paid services), 31 (7.8%) were using only voice calls (combination of free and paid services), and 31 (7.8%) were using only voice calls (combination of free and paid services).
Grover et al [71]	Medical colleges, government-funded institutes mental hospital setting, general hospital psychiatry units	India	Around 25% of institutes began offering telemental health services; 45.9% of institutes reported that telecommunication services continued during lockdown. Mental health services were being provided to people in quarantine (66.1%) and those with COVID-19 infection (59.6%), family members of patients with COVID-19, and those in quarantine (40.4%).
Humer et al [75]	Psychology/psychotherapy/counseling service	Czech Republic, Germany, Slovakia	Among all countries, the combined (personal contact + telephone + internet) number of patients treated on average per week during COVID-19 (mean 18.32 [SD 12.86]) did not differ from the combined (personal contact + telephone + internet) number of patients treated on average per week in the months before the COVID-19 situation (mean 19.35 [SD 13.73]), $t_{337}=-1.506$; $P=.133$.
Humer et al [75]	Psychology/psychotherapy/counseling service	Austria	During the COVID-19 pandemic, face-to-face psychotherapy remained the most abundant treatment modality.
Khanna et al [79]	CMHT and outpatient services	Australia	There was a 3% increase in appointment bookings compared with the same period in 2019. Cancellation/nonattendance rate dropped from an average of 19% last year to 12% for 2020
Kopec et al [80]	CMHT and outpatient services	United States	Prior to COVID-19, Network180 served an average of 2390 patients/month, which decreased to an average of 1921 patients/month during the pandemic. This decrease was noted most significantly in crisis services (averaging 822 patients/month before COVID-19 and 640 patients/month during COVID-19). Telehealth increased from 5% of all services prior to COVID-19 to 84% of all services during COVID-19. The majority of services provided via telehealth were audio only (versus audiovisual), with a ratio of 1.9:1 for crisis services and 4:1 for noncrisis services.

Author	Type of service	Service location	Adoption/coverage data
Looi et al [84]	Psychiatrist telehealth service	Australia	Percentage of consultations conducted using telemental health—ACT: 62% (April), 58% (May); NT: 53% (April), 51% (May); SA: 69% (April), 58% (May); Tasmania: 38% (April), 40% (May)
Looi et al [85]	Psychiatrist telehealth service	Australia	The majority of private practice was conducted by telehealth in April but was lower in May as new COVID-19 case rates fell. Percentage of consultations conducted using telemental health—NSW: 56% (April), 52% (May); QLD: 63% (April), 53% (May); VIC: 61% (April), 59% (May); WA: 51% (April), 36% (May)
Lynch et al [86]	CMHT and outpatient services	United States	The service continued providing all services except community-based coaching via telehealth. 90% of patients with complex psychosis accepted telehealth sessions and maintained their specific treatment plans. 2 opted out of telemental health. Mean comparisons between session attendance and cancellations/no-shows during the 6-week period before and after telehealth conversion showed no significant differences in service utilization.
Medalia et al [88]	CMHT and outpatient services	United States	Tracking the number of RS enrollees with active participation indicated that in the week before telehealth conversion, when shelter-in-place recommendations commenced, participation dropped from 94% to 52%; after telehealth conversion, participation rose from 67% in the first 4 days to 79% after 1 week and to 84% after 2 weeks.
Miu et al [89]	CMHT and outpatient services	United States	A total of 816 participants comprised the analytic sample. A total of 400 converted to telehealth and of those 64 were SMI[c]. The conversion rates from in-person psychotherapy to teletherapy were similar for SMI (n=64, 51.6%) and non-SMI (n=334, 48.3%) groups. The rate at which the SMI group converted from in-person therapy to teletherapy (52%) was not statistically different from that of the non-SMI group (48%) during COVID-19.
Patel et al [91]	All National Health Service Trust services	United Kingdom	From March 2020, in - person contacts reduced substantially from around 9000 per week to 3000 per week in early April 2020. Over the same period there was an increase in remote contacts from around 2500 per week in early March 2020 to around 8000 per week by the end of April 2020. Total clinical contacts per week dropped from around 12,500 in mid - March to around 10,000 in mid - April 2020. The number of unattended appointments was temporarily reduced in April, May, June, and September 2020.
Pierce et al [93]	Variety of MH service settings	United States	Psychologists estimated that telepsychology comprised 85.53% of their clinical work during the pandemic, compared with the prepandemic context when only 7.07% of their clinical work was conducted remotely.
Probst et al [94]	Psychology/psychotherapy/counseling service	Austria	Face-to-face psychotherapies in personal contact were reduced and remote psychotherapies (via telephone or internet) were increased in the early weeks of the COVID-19 lockdown as compared with the months before. Although average increases in psychotherapies via telephone (979%) or via internet (1561%) were dramatic, there was an undersupply of psychotherapy in Austria in the early weeks of the COVID-19 lockdown as the total number of patients treated on average per week was lower in COVID-19 lockdown than in the months before.
Reilly et al [95]	Various service types	United States	There was uptake of telemental health by approximately 80% of respondents by late March or early April 2020. All but 2.11% (19/903) of providers in this study made practice adjustments (transition to telemental health).
Rosen et al [98]	VHA[d] mental health services	United States	VHA provided nearly 1.2 million telephone and video encounters to veterans in April 2020 and reduced in-person visits by approximately 80% when compared with the October 2019 to February 2020 period before the pandemic. By June 2020, VHA had an 11-fold increase in encounters using direct-to-home video and a fivefold increase in telephone contacts relative to before the pandemic. VHA reduced in-person visits by approximately 80% when compared with the October 2019 to February 2020 period before the pandemic.

Author	Type of service	Service location	Adoption/coverage data
Scharff et al [100]	Community-based training clinic providing therapy	United States	The Psychological Services Centre saw an initial retention rate of 82% in the first week of teletherapy, with more clients resuming services in the weeks that followed.
Schlegl et al [101]	Inpatient mental health service	Germany	More than 80% of patients with bulimia nervosa received face-to-face therapy before the COVID-19 pandemic (81.8%) compared with 36.4% during the pandemic (ie, a decrease by 55.5%). Use of videoconference-based therapy increased from 3.6% to 21.8% and use of telephone contacts from 18.2% to 38.2%, whereas the use of additional online interventions decreased from 3.6% to 0%. Face-to-face psychotherapy decreased by 56% but videoconferencing therapy was only used by 22% of patients.
Sciarrino et al [102]	Veterans Healthcare Administration	United States	Approximately 76% of veterans engaged in posttraumatic stress disorder treatment chose to continue despite the COVID-19 pandemic via telehealth.
Sequeira et al [103]	Residential services	United States	After transition to teletherapy, the average daily virtual program census from March 19, 2020, to April 18, 2020, was 3.3 intensive outpatient program patients and 22.4 outpatients. These numbers indicate a slight decrease in intensive outpatient program patients (−0.3/per day) and an increase in outpatients seen per day (+2.7/per day). There was an increase in the outpatient sessions retained and a decrease in the appointments cancelled.
Severe et al [104]	CMHT and outpatient services	United States	Take up for remote care was over 95%; 82.8% (n=202) initially chose to receive psychiatric care through video visits, whereas 13.5% (n=33) chose telephone visits; 1.2% (n=3) decided to postpone care until in-person visit availability.
Sharma et al [28]	General hospital/physical health service	United States	By March 20, 2020, 67% of all outpatient appointments were conducted at home. Most of these appointments were conducted by phone with some TMH sessions. By March 27, 2020, 90% of all outpatient appointments were done at home, predominantly by phone (59%) but increasingly by HB-TMH[e] (31%). One week later (April 3, 2020), these rates were 48% versus 45%, respectively. By March 31, 2020, 98% of faculty completed expedited training and obtained departmental approval for HB-TMH services during the COVID-19 crisis. By April 10, 2020, HB-TMH was offered to all established outpatients for individual visits. Only the crisis clinic continued a regular in-clinic presence.
Sheehan et al [105]	CMHT and outpatient services	United Kingdom	64% were spending at least some time working from their workplace (either solely or in combination with home working); 33.9% were working from home only. Just over a third were at the workplace (n=178, 35.1%) and the remainder (n=147, 28.9%) worked from both home and at the workplace.
Termorshuizen et al [107]	CMHT and outpatient services	The Netherlands and United States	Most transitioned to telehealth care (United States 45%; Netherlands 42%), with fewer still receiving face - to - face care (United States 3%; Netherlands 6%), or not having been able to engage with their provider at all (United States 6%; Netherlands 5%).
Uscher-Pines et al [108]	CMHT and outpatient services and private hospital/clinic	United States	Most of the psychiatrists had transitioned to fully virtual practices. Only a quarter of the participants were seeing any patients in person.
Uscher-Pines et al [109]	CMHT and outpatient services; private hospital/clinic; general hospital/physical health service	United States	Telemedicine use: None (in-person only), 1 (5.6%); phone only, 2 (16.7%); video only, 0 (0.0%); combination of video and phone, 15 (83.3%).
van Dijk et al [110]	CMHT and outpatient services; psychology/psychotherapy/counseling service	The Netherlands	Treatment adherence was 100%.
Yellowlees et al [115]	General hospital/physical health service	United States	By the second day after shutdown, only 8% (n=52) of our appointments were in-person clinic visits, compared with our baseline average of 98%. By the third business day, 100% (n=73) of appointments were conducted virtually, with 92% (n=67) via videoconference and 8% (n=6) by phone.

Author	Type of service	Service location	Adoption/coverage data
Zulfic et al [116]	CMHT and outpatient services	Australia	Some patients still required regular face-to-face reviews, including the 91 patients (29%) who are treated with depot medications and 71 (23%) taking clozapine.

[a]CMHT: community mental health team.

[b]TMH-V: telemental health: video.

[c]SMI: serious mental illness.

[d]VHA: Veterans Health Administration.

[e]HB-TMH: home-based telemental health.

Appropriateness

There were some concerns raised over the appropriateness of remote care, for example, studies reported difficulties managing medication (eg, [45,108,109]) and concerns around engaging and assessing new patients (eg, [78]). Clinicians also found it harder to assess some physical indicators of mental health status (eg, hygiene, eye contact, physical symptoms of opioid withdrawal) without being able to see the service user in person. However, by contrast, remote methods of working felt safer for clinicians who worked with service users at risk for violence and behavioral dysregulation (eg, [56]). Online care was also not necessarily appropriate for patients with auditory or visual impairments, or with conditions such as migraines.

Staff reported concerns around the management of risk and safeguarding of service users when using remote methods of care (eg, [79,93]). Some helpful features of platforms which were thought to improve safety were using the waiting room function, being able to remove call participants, renaming participants (to protect anonymity), and using the private chat function.

Feasibility

In general, all studies reported good feasibility, at least for the short-term emergency response during the pandemic. However, some studies reported that telemental health is not suitable for all types of therapy, for example, those that require a physical presence (exposure therapy, role play, collaborative models; eg, [106]). Telemental health may also be less suitable for treating trauma [62,102]; clients with severe anxiety [62], learning difficulties or autism [105]; children [62]; and clients with cognitive impairment [62,90].

Cost-Effectiveness

There was limited information about costs of implementation of remote care in the included studies and no actual costs of telemental health were reported in the papers. However, initial evidence suggests remote care is not a costly intervention, with 1 paper stating that telemental health is “cost-effective” [69], while another mentions the use of “low-cost technologies” by clinicians [72].

Penetration

There was widespread penetration (the extent to which telemental health was integrated into mental health services) of remote methods of care delivery due to the COVID-19 pandemic, despite few services utilizing telemental health previously. Services were able to rapidly adapt to this new way of working, with the majority of appointments conducted remotely after the first few weeks of “stay at home” orders.

Sustainability

The sustainability of telemental health cannot be completely determined from the included studies, as they present data mostly from the early stages of the COVID-19 pandemic. However, there was some indication that although remote working was widely accepted as a necessity, once restrictions loosened, rates of telemental health use declined (corresponding with the drop in cases in Europe in summer 2020). This correlates with findings that not all staff and service users would want to continue using remote methods of care after the pandemic ends. However, there are some aspects of remote working that both clinicians and service users would like to keep in the future in combination with face-to-face care (eg, [93,104]), as this approach has benefits such as being more efficient, flexible, and enabling access for certain groups (eg, [7,60,62]).

Clinical Outcomes

Comparing the clinical outcomes of face-to-face and remote care using quantitative measures indicated that telemental health approaches could be as effective as face-to-face care (eg, [55,103]), although it should be noted that most studies were on a small scale. Several studies also reported no psychiatric decompensations after switching to remote care (eg, [86,88]). However, it is important to note that clinical outcomes for telemental health were not comparable for all service users; for example, Dores et al [60] found that a quarter of psychiatrists reported poorer clinical outcomes after switching to remote care. Another study also indicated that only one-third of clinicians felt as though telemental health consultations were comparable to prepandemic sessions [114]. A full presentation of the clinical outcomes reported in included studies is shown in Table 6.

Although the quality of therapeutic relationships reported by studies was generally good, clinicians reported problems reading patients’ emotions (eg, [43]) or feeling less connected to the service user compared with face-to-face sessions (eg, [56]). Clinicians also reported difficulties regarding feeling and expressing empathy remotely. Other challenges to therapeutic relationships when using remote care included a lack of client engagement, possible misunderstandings due to lack of nonverbal signals, common context, or not having a clear idea of patients’ physical state (alongside reduced privacy).

Table 6. Studies which reported clinical outcomes of telemental health.

Item	Type of service	Country	Clinical outcomes
Cheli et al [55]	Psychology/psychotherapy/counseling service	Italy	• 5/6 patients reported a reliable change index (≥1.96) in the primary outcome (Symptoms Checklist 90 [SCL-90] total score), and 1 reported a stable symptomatology. • All the patients reported a significant decreasing trend in the Depression, Anxiety and Stress Scale (DASS-21) total score (secondary outcome), as determined by Kendall τ (P<.001).
Dores et al [60]	Psychology/psychotherapy/counseling service	Portugal	• Comparing remote to in-person care (psychologists): 65 (71.6%) considered the results to be more of less the same, 4 (4.4%) reported obtaining better results with at-distance sessions, and 22 (24.2%) considered that at-distance sessions have yielded worse results than in-presence sessions. • Comparing remote with in-person care (service users): Remote and in-person sessions were more or less the same (n=71; 78.0%). Six (6.6%) of the respondents reported receiving better feedback (ie, the clients preferred the online sessions), and 1 (1.1%) received much better feedback. Even so, 13 (14.3%) psychologists received worse feedback from their clients about this type of intervention.
Erekson et al [61]	Psychology/psychotherapy/counseling service	United States	• Comparing current students (who received telemental health) with those in previous years (who received face-to-face care) found that students in previous years were not significantly different in their achievement of reliable improvement compared with those in 2020 (χ23=10.43, P=.015). • However, students in previous years were significantly more likely to deteriorate than those in 2020 (χ23=8.48, P=.04).
Gomet et al [67]	General hospital/physical health service (addictions service)	France	• 13 out of the 16 patients did not relapse during the data collection period.
Lai et al [81]	Day center (dementia service)	Hong Kong	*Patient outcomes* • The MoCA[a] scores in the intervention group (who received additional services using video conference, rather than telephone only) remained largely stable, whereas the MoCA scores for the control group fell after the 4-week study period ($F_{1,58}$=17.97, P<.001, ηp^2=0.24). • Quality of life scores were higher for the intervention group by the end of the study period ($F_{1,58}$=5.54, P<.05, np^2=0.49). • Scores on behavioral and psychological problems remained stable for both groups. *Caregiver outcomes* • Improvement in both physical and mental status of the caregivers was identified—($F_{1,58}$=60.30, P<.001, np^2=0.51) and ($F_{1,58}$=49.13, P<.001, np^2=0.46), respectively—a reduction in perceived burden ($F_{1,58}$=19.04, P<.001, np^2=0.25), and an increase in self-efficacy ($F_{1,58}$=17.30, P<.001, ηp^2=0.23).
Lynch et al [86]	CMHT[b] and outpatient services	United States	• During the 12-week study timeframe, the subsample of participants with complex psychosis remained psychiatrically stable; there were no psychiatric decompensations or referrals to a higher level of care.
Medalia et al [88]	CMHT and outpatient services	United States	• There were no psychiatric decompensations after conversion to telehealth.
Sequeira et al [103]	Residential services (obsessive compulsive disorder)	United States	• There were overall trends in reductions of scores of the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS), The Centre for Epidemiologic Studies Depression Scale (CES-D), The 7-item General Anxiety Disorder-7 (GAD-7), and Distress Intolerance Index (DII) across all patients, indicating that the telemental health program was effective in reducing symptoms of obsessive compulsive disorder, anxiety, and depression.

Item	Type of service	Country	Clinical outcomes
Wyler et al [114]	Mental health team and outpatient services	Switzerland	• For about 1 in 3 cases, therapists reported that they felt the sessions were at least fairly comparable to pre-COVID-19 sessions or that the restrictions were not particularly problematic.

[a]MoCA: Montreal Cognitive Assessment.

[b]CMHT: community mental health team.

Social Outcomes

One study [81] compared social outcomes in a trial comparing telephone-only care with caregivers of older adults with neurocognitive disorder, with supplementary video care to both carers and service users. Findings indicated that those who received both telephone and video support had greater resilience, better cognitive functioning, and a higher quality of life.

Organizational and Care Delivery Outcomes

Improved communication was noted between staff when using telemental health when compared with traditional face-to-face care, as the use of online file sharing or discussion platforms facilitated communication between staff (eg, [74,105]). The use of online methods also facilitated staff training and some staff reported that remote working resulted in a better work–life balance (eg, [59]).

Discussion

Summary of Findings

This review collated evidence regarding the implementation and outcomes of remote working in mental health services in the context of the COVID-19 pandemic. Most studies indicate a relatively high level of activity, suggesting that at least in the services studied in higher-income countries, much mental health care can be shifted to telemental health in a crisis. Services mainly reported using a mixture of phone and video calls, with both service users and clinicians varying in their preference for these modalities. There were some indicators of reduced numbers of missed appointments, potentially due to the greater convenience of remote care, which may make access to services easier for some service users.

There was reasonable acceptability across the studies, at least in conditions where the alternative may have been no contact with services at all. However, there were situations where telemental health may be less acceptable, including for new patients, physical health aspects of care, and for service users without a private space at home to use for therapy. Telemental health also may not be as feasible for certain types of support, including support which needs a physical presence such as exposure therapy or role play. This finding reflects those in a systematic review by Turgoose et al [117], which found that service users had concerns around managing emotions during exposure tasks without the physical presence of a clinician. There was also evidence that telemental health may not be feasible for some clinical presentations, including some service users with psychosis, learning difficulties, or autism. Clinicians also reported a decrease in their ability to develop and maintain a strong therapeutic relationship with service users, due to being unable to pick up on nonverbal cues and a lack of connectedness, something which was also identified in a review conducted prior to the pandemic [118]. The acceptability levels found in this study are not dissimilar to previous studies (eg, [15]), even though the participants in the current studies are less likely to be volunteering to pilot a new type of care and more likely to be using telemental health because they have no alternative.

Few formal investigations of how to improve implementation were identified in this review, which may reflect the rapid nature of research conducted during the pandemic. However, some strategies for improving adoption/penetration/acceptability may include staff training, the use of telemental health champions, strategies for introducing service users to technology, and providing some simple guidance on how to use it best, identifying situations or populations when telemental health is not a good idea and those where it might be better. There was also a lack of fidelity assessments when therapies had to be adapted to fit telemental health delivery formats; therefore, little is known about the consequences of these adaptations.

Our interpretation of this pattern of findings is that the successful delivery in a pandemic of telemental health should not necessarily be seen as confirmation that people are happy with this mode of delivery long-term, as some of the identified problems may become more serious over time, and reports of being satisfied may have reflected awareness that at the time of the study, it was difficult to offer care by any other means. The longevity of these changes will ultimately turn not only on information technology, safety, and quality, but also on whether policy changes will support the reimbursements and regulatory adjustments implemented during the current crisis [29,57].

Implications for Future Research

There was a lack of reporting in included studies of trying to identify and reach those patients who are at increased risk of digital exclusion (Textbox 1). The needs of those at risk of digital exclusion are still largely underreported in both pre-COVID-19 [15] and COVID-19–specific literature and should be made a priority for future research. Studies also included little information regarding the cost-effectiveness of telemental health implementation. Further research is needed to explore the differences in cost (both to the service and to the service user) between face-to-face and telemental health care. Further research can also formally compare (rather than simply observe) different delivery support strategies that can improve the implementation and potentially also the clinical effectiveness of telemental health, including for specific conditions and service user groups.

Textbox 1. Lived experience perspective commentary.

Lived experience commentary by Karen Machin and Raza Griffiths, members of the NIHR Mental Health Policy Unit's Lived Experience Working Group.

Systematic reviews aim to give an overview of research findings around a particular topic, although, as in this review, they may find that many of the primary studies are of moderate to low quality. At times of intense pressure, such as those presented by the COVID pandemic, overviews may be welcomed, and researchers may feel a need to respond promptly. This urgency creates a potential for gaps to be overlooked.

The main concern for us is around service user involvement, which has been acknowledged as key to good quality research, but has still not become the default for research teams. Participatory approaches are largely absent from the reviewed studies, with service user and carer views "reported by clinicians" ignoring the likelihood of misunderstandings and service user views being skewed by unequal power dynamics. There is no statement about the involvement of lived experience researchers in the review itself either. While some researchers may not wish to expose their personal experiences, it is important to be clear about the level of involvement within all studies.

Service users can shine a spotlight on aspects of mental health support of importance to people who rely on services which may otherwise be missed. Their lack of involvement in this review may be one reason why only one out of 41 primary research studies reviewed looks at the voluntary sector, and none at peer support or social care. It is unclear if this is a result of the review process itself or a lack of evidence. Had service users been involved, this fundamental gap might have been explored at an earlier stage.

The review does identify some factors that can help determine how useful or accessible telemental health is for different groups of people. However, an analysis linking these factors to broader underlying factors such as poverty would be more helpful.

Further research is needed to ensure future service planning can more accurately assess which elements of telemental health work well, why, and for whom, and to have service user involvement integral to the whole process from beginning to end.

Finally, there is scope to conduct big data studies to identify who is not accessing remote care or those at risk of disengaging, and potential comparisons for matched groups to try to compare effectiveness across a range of settings, as this could be done more quickly than in clinical trials while respecting patient preference.

Strengths and Limitations

The studies included in this review identified outcomes across different settings and health care systems, which may help findings generalize to different settings. This review also captured recent findings on the use of telemental health during the COVID-19 pandemic, allowing findings to be used to improve both existing and future models of remote mental health care.

However, it is also important to take some limitations into account when interpreting the findings from this review. First, the results from quality assessment indicated that while around half of primary research studies and the majority of the service evaluations were high quality, around half of primary research studies were scored as moderate to low. This reflects the short nature of studies and often quick turnaround from data collection to publication. Some studies were also published in preprint form and therefore had not undergone peer review. The majority of studies used cross-sectional data, rather than more rigorous methods. Second, there was a lack of high-quality quantitative evidence for the clinical effectiveness of telemental health care. Clinical effectiveness outcomes were only reported in 9/77 included studies, with some of these findings only based on qualitative evidence or a small number of service users. It is also important to note that the voices of those who dropped out of care may not be included.

The short time scale for data collection and assessment of changes in practice in included studies could also be viewed as a limitation of this research, as it is not clear if changes will be sustained over time or in other contexts (eg, lower-income countries). We also recognize that the search dates do not cover the whole of the pandemic to date; however, as plans for the medium- to long-term adoption of telemental health are currently being made in several countries, we think it best to make the findings from this research available promptly. Research was also not inclusive of those not accessing or using remote technologies, meaning there is a risk of those at risk of digital exclusion being forgotten when taking the findings of this review into consideration.

We also designed this review to be conducted rapidly to ensure results would be relevant and quickly available, therefore we chose to search 4 databases and not all studies were independently double screened by blinded researchers. Although quality assessment was conducted by 2 reviewers independently, they were also not blinded to the previous decision. However, we are confident that the rigor of our searches and inclusion of preprint servers meant that the papers included are representative of the literature on this topic.

Conclusion

Telemental health was a largely effective method to enable continuation of mental health support during the COVID-19 pandemic. While most reported outcomes were positive, telemental health was not feasible for all types of support and may not be acceptable to all service user groups. A blended approach combining face-to-face and telemental health care may be the most desirable service model for future care. The need remains for higher-quality evidence regarding the clinical effectiveness of telemental health and how uptake can be improved for groups at risk of digital exclusion.

Acknowledgments

Policy Research Unit

This paper presents independent research commissioned and funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted by the NIHR Policy Research Unit (PRU) in Mental Health. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care or its arm's length bodies, or other government departments.

Applied Research Collaboration South London and King's Improvement Science

This study is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. NS and LG are members of King's Improvement Science, which offers co-funding to the NIHR ARC South London and is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), and Guy's and St Thomas' Foundation. The views expressed are those of the author(s) and not necessarily those of the NIHR, the NHS, the charities, or the Department of Health and Social Care.

LG is funded by King's Improvement Science, which is part of the NIHR ARC South London and comprises a specialist team of improvement scientists and senior researchers based at King's College London. King's Improvement Science is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), and the Guy's and St Thomas' Foundation.

Conflicts of Interest

NS is the director of the London Safety and Training Solutions Ltd, which offers training in patient safety, implementation solutions and human factors to healthcare organizations and the pharmaceutical industry. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Supplementary Table 1 - PubMed search strategy.
[DOCX File , 14 KB - jmir_v23i12e31746_app1.docx]

Multimedia Appendix 2
Supplementary Table 2 - Details of study characteristics.
[DOCX File , 45 KB - jmir_v23i12e31746_app2.docx]

References

1. World Health Organization. The Impact of COVID-19 on Mental, Neurological and Substance Use Services: Results of a Rapid Assessment. 2020. URL: https://www.who.int/publications/i/item/978924012455 [accessed 2021-11-29]
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. 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]
4. Sheridan Rains L, Johnson S, Barnett P, Steare T, Needle J, Carr S, COVID-19 Mental Health Policy Research Unit Group. Early impacts of the COVID-19 pandemic on mental health care and on people with mental health conditions: framework synthesis of international experiences and responses. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):13-24 [FREE Full text] [doi: 10.1007/s00127-020-01924-7] [Medline: 32804258]
5. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
6. American Psychological Association. What Are Telehealth and Telepsychology. 2017. URL: https://www.apa.org/pi/disability/resources/publications/telepsychology [accessed 2021-11-25]
7. Johnson S, Dalton-Locke C, Vera San Juan N, Foye U, Oram S, Papamichail A, COVID-19 Mental Health Policy Research Unit Group. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):25-37 [FREE Full text] [doi: 10.1007/s00127-020-01927-4] [Medline: 32857218]
8. Telehealth bibliography. Telebehavioural Health Institute. URL: https://telehealth.org/bibliography/ [accessed 2021-12-01]
9. Bashshur RL, Shannon GW, Bashshur N, Yellowlees PM. The Empirical Evidence for Telemedicine Interventions in Mental Disorders. Telemed J E Health 2016 Mar;22(2):87-113 [FREE Full text] [doi: 10.1089/tmj.2015.0206] [Medline: 26624248]

10. Salmoiraghi A, Hussain S. A Systematic Review of the Use of Telepsychiatry in Acute Settings. J Psychiatr Pract 2015 Sep;21(5):389-393. [doi: 10.1097/PRA.0000000000000103] [Medline: 26348806]
11. Hilty DM, Ferrer DC, Parish MB, Johnston B, Callahan EJ, Yellowlees PM. The effectiveness of telemental health: a 2013 review. Telemed J E Health 2013 Jun;19(6):444-454 [FREE Full text] [doi: 10.1089/tmj.2013.0075] [Medline: 23697504]
12. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
13. Varker T, Brand RM, Ward J, Terhaag S, Phelps A. Efficacy of synchronous telepsychology interventions for people with anxiety, depression, posttraumatic stress disorder, and adjustment disorder: A rapid evidence assessment. Psychol Serv 2019 Nov;16(4):621-635. [doi: 10.1037/ser0000239] [Medline: 29809025]
14. Fletcher TL, Hogan JB, Keegan F, Davis ML, Wassef M, Day S, et al. Recent Advances in Delivering Mental Health Treatment via Video to Home. Curr Psychiatry Rep 2018 Jul 21;20(8):56. [doi: 10.1007/s11920-018-0922-y] [Medline: 30032337]
15. Barnett P, Goulding L, Casetta C, Jordan H, Sheridan-Rains L, Steare T, et al. Implementation of Telemental Health Services Before COVID-19: Rapid Umbrella Review of Systematic Reviews. J Med Internet Res 2021 Jul 20;23(7):e26492 [FREE Full text] [doi: 10.2196/26492] [Medline: 34061758]
16. Gloff NE, LeNoue SR, Novins DK, Myers K. Telemental health for children and adolescents. Int Rev Psychiatry 2015 Nov 05;27(6):513-524. [doi: 10.3109/09540261.2015.1086322] [Medline: 26540584]
17. Seritan AL, Heiry M, Iosif A, Dodge M, Ostrem JL. Telepsychiatry for patients with movement disorders: a feasibility and patient satisfaction study. J Clin Mov Disord 2019;6:1 [FREE Full text] [doi: 10.1186/s40734-019-0077-y] [Medline: 31183157]
18. Santesteban-Echarri O, Piskulic D, Nyman RK, Addington J. Telehealth interventions for schizophrenia-spectrum disorders and clinical high-risk for psychosis individuals: A scoping review. J Telemed Telecare 2020;26(1-2):14-20. [doi: 10.1177/1357633X18794100] [Medline: 30134781]
19. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: A systematic review. J Telemed Telecare 2019 Jul;25(6):327-342. [doi: 10.1177/1357633X18775855] [Medline: 29788807]
20. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clinical Psychology: Science and Practice 2020 Jun;27(2):e12311. [doi: 10.1111/cpsp.12311]
21. Vera San Juan N, Shah P, Schlief M, Appleton R, Nyikavaranda P, Birken M, et al. Service user experiences and views regarding telemental health during the COVID-19 pandemic: A co-produced framework analysis. PLoS One 2021;16(9):e0257270 [FREE Full text] [doi: 10.1371/journal.pone.0257270] [Medline: 34529705]
22. Stoll J, Müller JA, Trachsel M. Ethical Issues in Online Psychotherapy: A Narrative Review. Front Psychiatry 2019 Feb 11;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
23. Tullio V, Perrone G, Bilotta C, Lanzarone A, Argo A. Psychological support and psychotherapy via digital devices in Covid-19 emergency time: Some critical issues. Med Leg J 2020 Jul 03;88(2):73-76. [doi: 10.1177/0025817220926942] [Medline: 32490713]
24. Europe Economics. Regulatory Approaches to Telemedicine. 2018. URL: https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/research-and-insight-archive/regulatory-approaches-to-telemedicine [accessed 2021-11-26]
25. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
26. Cowan K, McKean A, Gentry M, Hilty D. Barriers to Use of Telepsychiatry: Clinicians as Gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
27. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving Implementation of eMental Health for Mood Disorders in Routine Practice: Systematic Review of Barriers and Facilitating Factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
28. Sharma A, Sasser T, Schoenfelder Gonzalez E, Vander Stoep A, Myers K. Implementation of Home-Based Telemental Health in a Large Child Psychiatry Department During the COVID-19 Crisis. J Child Adolesc Psychopharmacol 2020 Sep 01;30(7):404-413. [doi: 10.1089/cap.2020.0062] [Medline: 32639849]
29. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid Increase in Telemental Health Within the Department of Veterans Affairs During the COVID-19 Pandemic. Telemed J E Health 2021 Apr 01;27(4):454-458. [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
30. Gentry MT, Puspitasari AJ, McKean AJ, Williams MD, Breitinger S, Geske JR, et al. Clinician Satisfaction with Rapid Adoption and Implementation of Telehealth Services During the COVID-19 Pandemic. Telemed J E Health 2021 Mar 19:e1 (forthcoming). [doi: 10.1089/tmj.2020.0575] [Medline: 33606560]
31. Nicholas J, Bell IH, Thompson A, Valentine L, Simsir P, Sheppard H, et al. Implementation lessons from the transition to telehealth during COVID-19: a survey of clinicians and young people from youth mental health services. Psychiatry Res 2021 May;299:113848. [doi: 10.1016/j.psychres.2021.113848] [Medline: 33725578]

32. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and Telepsychiatry: Development of Evidence-Based Guidance for Clinicians. JMIR Ment Health 2020 Aug 28;7(8):e21108 [FREE Full text] [doi: 10.2196/21108] [Medline: 32658857]
33. 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]
34. The Endnote Team. Endnote X9. Philadelphia, PA: Clarivate; 2013.
35. 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]
36. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
37. 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]
38. Safaeinili N, Brown-Johnson C, Shaw JG, Mahoney M, Winget M. CFIR simplified: Pragmatic application of and adaptations to the Consolidated Framework for Implementation Research (CFIR) for evaluation of a patient-centered care transformation within a learning health system. Learn Health Syst 2020 Sep 26;4(1):e10201 [FREE Full text] [doi: 10.1002/lrh2.10201] [Medline: 31989028]
39. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M. EPPI-Reviewer: Advanced Software for Systematic Reviews, Maps and Evidence Synthesis. London, UK: UCL Social Research Institute; 2020.
40. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
41. Tyndall J. AACODS Checklist. 2010. URL: https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/AACODS_Checklist.pdf [accessed 2021-01-12]
42. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2021-11-25]
43. Aafjes-van Doorn K, Békés V, Prout TA. Grappling with our therapeutic relationship and professional self-doubt during COVID-19: will we use video therapy again? Counselling Psychology Quarterly 2020 Jun 03:1-12. [doi: 10.1080/09515070.2020.1773404]
44. Anton MT, Ridings LE, Gavrilova Y, Bravoco O, Ruggiero KJ, Davidson TM. Transitioning a technology-assisted stepped-care model for traumatic injury patients to a fully remote model in the age of COVID-19. Counselling Psychology Quarterly 2020 Jun 26:1-12. [doi: 10.1080/09515070.2020.1785393]
45. 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]
46. Békés V, Aafjes-van Doorn K. Psychotherapists' attitudes toward online therapy during the COVID-19 pandemic. Journal of Psychotherapy Integration 2020 Jun;30(2):238-247. [doi: 10.1037/int0000214]
47. Békés V, Aafjes-van Doorn K, Prout TA, Hoffman L. Stretching the Analytic Frame: Analytic Therapists' Experiences with Remote Therapy During COVID-19. J Am Psychoanal Assoc 2020 Jun 26;68(3):437-446. [doi: 10.1177/0003065120939298] [Medline: 32589042]
48. Benaque A, Gurruchaga MJ, Abdelnour C, Hernández I, Cañabate P, Alegret M, Research CenterMemory Clinic, Fundació ACE. Dementia Care in Times of COVID-19: Experience at Fundació ACE in Barcelona, Spain. J Alzheimers Dis 2020;76(1):33-40 [FREE Full text] [doi: 10.3233/JAD-200547] [Medline: 32538856]
49. Berdullas Saunders S, Gesteira Santos C, Morán Rodríguez N, Fernández Hermida JR, Santolaya F, Sanz Fernández J, et al. [The COVID-19 psychological helpline of the Spanish Ministry of Health and Spanish Psychological Association: characteristics and demand.]. Rev Esp Salud Publica 2020 Oct 14;94:e1 (forthcoming). [Medline: 33056957]
50. Bhome R, Huntley J, Dalton-Locke C, San Juan NV, Oram S, Foye U, et al. Impact of the COVID-19 pandemic on older adults mental health services: A mixed methods study. Int J Geriatr Psychiatry 2021 Nov;36(11):1748-1758 [FREE Full text] [doi: 10.1002/gps.5596] [Medline: 34216045]
51. Bierbooms JJPA, van Haaren M, IJsselsteijn WA, de Kort YAW, Feijt M, Bongers IMB. Integration of Online Treatment Into the "New Normal" in Mental Health Care in Post-COVID-19 Times: Exploratory Qualitative Study. JMIR Form Res 2020 Oct 08;4(10):e21344 [FREE Full text] [doi: 10.2196/21344] [Medline: 33001835]
52. Boldrini T, Schiano Lomoriello A, Del Corno F, Lingiardi V, Salcuni S. Psychotherapy During COVID-19: How the Clinical Practice of Italian Psychotherapists Changed During the Pandemic. Front Psychol 2020 Oct 21;11:591170 [FREE Full text] [doi: 10.3389/fpsyg.2020.591170] [Medline: 33192932]
53. Burton A, McKinlay A, Aughterson H, Fancourt D. Impact of the COVID-19 pandemic on the mental health and well-being of adults with mental health conditions in the UK: a qualitative interview study. J Ment Health 2021 Jul 29:1-8. [doi: 10.1080/09638237.2021.1952953] [Medline: 34323635]

54. Carpiniello B, Tusconi M, di Sciascio G, Zanalda E, di Giannantonio M, Executive Committee of the Italian Society of Psychiatry. Mental health services in Italy during the COVID-19 pandemic. Psychiatry Clin Neurosci 2020 Aug 30;74(8):442-443 [FREE Full text] [doi: 10.1111/pcn.13082] [Medline: 32515525]
55. Cheli S, Cavalletti V, Petrocchi N. An online compassion-focused crisis intervention during COVID-19 lockdown: a cases series on patients at high risk for psychosis. Psychosis 2020 Jul 20;12(4):359-362. [doi: 10.1080/17522439.2020.1786148]
56. Chen JA, Chung W, Young SK, Tuttle MC, Collins MB, Darghouth SL, et al. COVID-19 and telepsychiatry: Early outpatient experiences and implications for the future. Gen Hosp Psychiatry 2020 Sep;66:89-95 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.07.002] [Medline: 32750604]
57. Childs AW, Unger A, Li L. Rapid design and deployment of intensive outpatient, group-based psychiatric care using telehealth during coronavirus disease 2019 (COVID-19). J Am Med Inform Assoc 2020 Jul 01;27(9):1420-1424 [FREE Full text] [doi: 10.1093/jamia/ocaa138] [Medline: 32687151]
58. Colle R, Ait Tayeb AEK, de Larminat D, Commery L, Boniface B, Lasica P, et al. Short-term acceptability by patients and psychiatrists of the turn to psychiatric teleconsultation in the context of the COVID-19 pandemic. Psychiatry Clin Neurosci 2020 Aug 28;74(8):443-444 [FREE Full text] [doi: 10.1111/pcn.13081] [Medline: 32511825]
59. Datta N, Derenne J, Sanders M, Lock JD. Telehealth transition in a comprehensive care unit for eating disorders: Challenges and long-term benefits. Int J Eat Disord 2020 Nov 27;53(11):1774-1779. [doi: 10.1002/eat.23348] [Medline: 32715512]
60. Dores AR, Geraldo A, Carvalho IP, Barbosa F. The Use of New Digital Information and Communication Technologies in Psychological Counseling during the COVID-19 Pandemic. Int J Environ Res Public Health 2020 Oct 21;17(20):7663 [FREE Full text] [doi: 10.3390/ijerph17207663] [Medline: 33096650]
61. Erekson DM, Bailey RJ, Cattani K, Fox ST, Goates-Jones MK. Responding to the Covid-19 pandemic at a university counseling center: administrative actions, client retention, and psychotherapy outcome. Counselling Psychology Quarterly 2020 Aug 17:1-15. [doi: 10.1080/09515070.2020.1807914]
62. 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 01;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
63. Fernández-Aranda F, Munguía L, Mestre-Bach G, Steward T, Etxandi M, Baenas I, et al. COVID Isolation Eating Scale (CIES): Analysis of the impact of confinement in eating disorders and obesity-A collaborative international study. Eur Eat Disord Rev 2020 Nov;28(6):871-883 [FREE Full text] [doi: 10.1002/erv.2784] [Medline: 32954595]
64. Foye U, Dalton-Locke C, Harju-Seppänen J, Lane R, Beames L, Vera San Juan N, et al. How has COVID-19 affected mental health nurses and the delivery of mental health nursing care in the UK? Results of a mixed-methods study. J Psychiatr Ment Health Nurs 2021 Apr;28(2):126-137 [FREE Full text] [doi: 10.1111/jpm.12745] [Medline: 33608956]
65. Gaddy S, Gallardo R, McCluskey S, Moore L, Peuser A, Rotert R. COVID-19 and Music Therapists' Employment, Service Delivery, Perceived Stress, and Hope: A Descriptive Study. Music Therapy Perspectives 2020 Sep 08;38(2):157-166. [doi: 10.1093/mtp/miaa018]
66. Gillard S, Dare C, Hardy J, Nyikavaranda P, Rowan Olive R, Shah P, NIHR Mental Health Policy Research Unit Covid coproduction research group. Experiences of living with mental health problems during the COVID-19 pandemic in the UK: a coproduced, participatory qualitative interview study. Soc Psychiatry Psychiatr Epidemiol 2021 Aug;56(8):1447-1457 [FREE Full text] [doi: 10.1007/s00127-021-02051-7] [Medline: 33665680]
67. Gomet R, Souchet C, Mouellic ML, Maafa M, Labey A. Addiction outpatient hospital reinvents itself during the Coronavirus outbreak. Alcoologie et Addictologie 2020;42(1):54-57.
68. Graell M, Morón-Nozaleda MG, Camarneiro R, Villaseñor Á, Yáñez S, Muñoz R, et al. Children and adolescents with eating disorders during COVID-19 confinement: Difficulties and future challenges. Eur Eat Disord Rev 2020 Nov 29;28(6):864-870. [doi: 10.1002/erv.2763] [Medline: 32729139]
69. Grover S, Mishra E, Chakrabarti S, Mehra A, Sahoo S. Telephonic monitoring of patients on clozapine in the resource-poor setting during the COVID-19 pandemic. Schizophr Res 2020 Aug;222:489-490 [FREE Full text] [doi: 10.1016/j.schres.2020.05.062] [Medline: 32563610]
70. Grover S, Mehra A, Sahoo S, Avasthi A, Tripathi A, D'Souza A, et al. Impact of COVID-19 pandemic and lockdown on the state of mental health services in the private sector in India. Indian J Psychiatry 2020;62(5):488. [doi: 10.4103/psychiatry.indianjpsychiatry_568_20]
71. Grover S, Mehra A, Sahoo S, Avasthi A, Tripathi A, D'Souza A, et al. State of mental health services in various training centers in India during the lockdown and COVID-19 pandemic. Indian J Psychiatry 2020;62(4):363. [doi: 10.4103/psychiatry.indianjpsychiatry_567_20]
72. Haxhihamza K, Arsova S, Bajraktarov S, Kalpak G, Stefanovski B, Novotni A, et al. Patient Satisfaction with Use of Telemedicine in University Clinic of Psychiatry: Skopje, North Macedonia During COVID-19 Pandemic. Telemed J E Health 2021 Apr 01;27(4):464-467. [doi: 10.1089/tmj.2020.0256] [Medline: 32809916]
73. He Z, Chen J, Pan K, Yue Y, Cheung T, Yuan Y, et al. The development of the 'COVID-19 Psychological Resilience Model' and its efficacy during the COVID-19 pandemic in China. Int J Biol Sci 2020;16(15):2828-2834 [FREE Full text] [doi: 10.7150/ijbs.50127] [Medline: 33061799]

74. Hom MA, Weiss RB, Millman ZB, Christensen K, Lewis EJ, Cho S, et al. Development of a virtual partial hospital program for an acute psychiatric population: Lessons learned and future directions for telepsychotherapy. Journal of Psychotherapy Integration 2020 Jun;30(2):366-382. [doi: 10.1037/int0000212]
75. Humer E, Pieh C, Kuska M, Barke A, Doering BK, Gossmann K, et al. Provision of Psychotherapy during the COVID-19 Pandemic among Czech, German and Slovak Psychotherapists. Int J Environ Res Public Health 2020 Jul 04;17(13):4811 [FREE Full text] [doi: 10.3390/ijerph17134811] [Medline: 32635422]
76. Humer E, Stippl P, Pieh C, Schimböck W, Probst T. Psychotherapy via the Internet: What Programs Do Psychotherapists Use, How Well-Informed Do They Feel, and What Are Their Wishes for Continuous Education? Int J Environ Res Public Health 2020 Nov 05;17(21):8182 [FREE Full text] [doi: 10.3390/ijerph17218182] [Medline: 33167478]
77. Izakova L, Breznoscakova D, Jandova K, Valkucakova V, Bezakova G, Suvada J. What mental health experts in Slovakia are learning from COVID-19 pandemic? Indian J Psychiatry 2020;62(9):459. [doi: 10.4103/psychiatry.indianjpsychiatry_758_20]
78. Jurcik T, Jarvis GE, Zeleskov Doric J, Krasavtseva Y, Yaltonskaya A, Ogiwara K, et al. Adapting mental health services to the COVID-19 pandemic: reflections from professionals in four countries. Counselling Psychology Quarterly 2020 Jul 16:1-27. [doi: 10.1080/09515070.2020.1785846]
79. Khanna R, Murnane T, Kumar S, Rolfe T, Dimitrieski S, McKeown M, et al. Making working from home work: reflections on adapting to change. Australas Psychiatry 2020 Oct 10;28(5):504-507. [doi: 10.1177/1039856220953701] [Medline: 32910690]
80. Kopec K, Janney CA, Johnson B, Spykerman K, Ryskamp B, Achtyes ED. Rapid Transition to Telehealth in a Community Mental Health Service Provider During the COVID-19 Pandemic. Prim. Care Companion CNS Disord 2020 Oct 01;22(5):e1. [doi: 10.4088/pcc.20br02787]
81. Lai FH, Yan EW, Yu KK, Tsui W, Chan DT, Yee BK. The Protective Impact of Telemedicine on Persons With Dementia and Their Caregivers During the COVID-19 Pandemic. Am J Geriatr Psychiatry 2020 Nov;28(11):1175-1184 [FREE Full text] [doi: 10.1016/j.jagp.2020.07.019] [Medline: 32873496]
82. Lakeman R, Crighton J. The Impact of Social Distancing on People with Borderline Personality Disorder: The Views of Dialectical Behavioural Therapists. Issues Ment Health Nurs 2021 May 15;42(5):410-416. [doi: 10.1080/01612840.2020.1817208] [Medline: 32931341]
83. Lin X, Swift J, Cheng Y, An Q, Liang H, Wang Y. The Psychological Hotline Services Quality Survey during the Pandemic of COVID-19 in Mainland China. International Journal of Mental Health Promotion 2020;22(3):109-113.
84. Looi JC, Allison S, Bastiampillai T, Pring W. Private practice metropolitan telepsychiatry in smaller Australian jurisdictions during the COVID-19 pandemic: preliminary analysis of the introduction of new Medicare Benefits Schedule items. Australas Psychiatry 2020 Dec 05;28(6):639-643. [doi: 10.1177/1039856220960381] [Medline: 33016772]
85. Looi JC, Allison S, Bastiampillai T, Pring W. Private practice metropolitan telepsychiatry in larger Australian states during the COVID-19 pandemic: an analysis of the first 2 months of new MBS telehealth item psychiatrist services. Australas Psychiatry 2020 Dec 30;28(6):644-648. [doi: 10.1177/1039856220961906] [Medline: 32997521]
86. Lynch DA, Medalia A, Saperstein A. The Design, Implementation, and Acceptability of a Telehealth Comprehensive Recovery Service for People With Complex Psychosis Living in NYC During the COVID-19 Crisis. Front Psychiatry 2020 Sep 8;11:581149 [FREE Full text] [doi: 10.3389/fpsyt.2020.581149] [Medline: 33101093]
87. McBeath AG, du Plock S, Bager-Charleson S. The challenges and experiences of psychotherapists working remotely during the coronavirus pandemic. Couns Psychother Res 2020 Jun 23;20(3):394-405 [FREE Full text] [doi: 10.1002/capr.12326] [Medline: 32837329]
88. Medalia A, Lynch DA, Herlands T. Telehealth Conversion of Serious Mental Illness Recovery Services During the COVID-19 Crisis. Psychiatr Serv 2020 Aug 01;71(8):872-872. [doi: 10.1176/appi.ps.71705] [Medline: 32741337]
89. Miu AS, Vo HT, Palka JM, Glowacki CR, Robinson RJ. Teletherapy with serious mental illness populations during COVID-19: telehealth conversion and engagement. Counselling Psychology Quarterly 2020 Jul 12:1-18. [doi: 10.1080/09515070.2020.1791800]
90. Olwill C, Mc Nally D, Douglas L. Psychiatrist experience of remote consultations by telephone in an outpatient psychiatric department during the COVID-19 pandemic. Ir J Psychol Med 2021 Jun 22;38(2):132-139 [FREE Full text] [doi: 10.1017/ipm.2020.51] [Medline: 32438945]
91. Patel R, Irving J, Brinn A, Broadbent M, Shetty H, Pritchard M, et al. Impact of the COVID-19 pandemic on remote mental healthcare and prescribing in psychiatry: an electronic health record study. BMJ Open 2021 Mar 30;11(3):e046365 [FREE Full text] [doi: 10.1136/bmjopen-2020-046365] [Medline: 33785494]
92. Peralta E, Taveras M. Effectiveness of teleconsultation use in access to mental health services during the coronavirus disease 2019 pandemic in the Dominican Republic. Indian J Psychiatry 2020;62(9):492. [doi: 10.4103/psychiatry.indianjpsychiatry_1047_20]
93. Pierce BS, Perrin PB, Tyler CM, McKee GB, Watson JD. The COVID-19 telepsychology revolution: A national study of pandemic-based changes in U.S. mental health care delivery. Am Psychol 2021 Jan;76(1):14-25. [doi: 10.1037/amp0000722] [Medline: 32816503]

94. Probst T, Stippl P, Pieh C. Changes in Provision of Psychotherapy in the Early Weeks of the COVID-19 Lockdown in Austria. Int J Environ Res Public Health 2020 May 27;17(11):3815 [FREE Full text] [doi: 10.3390/ijerph17113815] [Medline: 32471295]
95. Reilly SE, Zane KL, McCuddy WT, Soulliard ZA, Scarisbrick DM, Miller LE, et al. Mental Health Practitioners' Immediate Practical Response During the COVID-19 Pandemic: Observational Questionnaire Study. JMIR Ment Health 2020 Oct 01;7(9):e21237 [FREE Full text] [doi: 10.2196/21237] [Medline: 32931440]
96. Roach P, Zwiers A, Cox E, Fischer K, Charlton A, Josephson C, et al. Understanding the impact of the COVID-19 pandemic on well-being and virtual care for people living with dementia and care partners living in the community. Dementia (London) 2021 Aug;20(6):2007-2023 [FREE Full text] [doi: 10.1177/1471301220977639] [Medline: 33381996]
97. Roncero C, García-Ullán L, de la Iglesia-Larrad JI, Martín C, Andrés P, Ojeda A, Montejo. The response of the mental health network of the Salamanca area to the COVID-19 pandemic: The role of the telemedicine. Psychiatry Res 2020 Sep;291:113252 [FREE Full text] [doi: 10.1016/j.psychres.2020.113252] [Medline: 32623263]
98. Rosen CS, Morland LA, Glassman LH, Marx BP, Weaver K, Smith CA, et al. Virtual mental health care in the Veterans Health Administration's immediate response to coronavirus disease-19. Am Psychol 2021 Jan;76(1):26-38. [doi: 10.1037/amp0000751] [Medline: 33119331]
99. Sasangohar F, Bradshaw MR, Carlson MM, Flack JN, Fowler JC, Freeland D, et al. Adapting an Outpatient Psychiatric Clinic to Telehealth During the COVID-19 Pandemic: A Practice Perspective. J Med Internet Res 2020 Oct 01;22(10):e22523 [FREE Full text] [doi: 10.2196/22523] [Medline: 32936768]
100. Scharff A, Breiner CE, Ueno LF, Underwood SB, Merritt EC, Welch LM, et al. Shifting a training clinic to teletherapy during the COVID-19 pandemic: a trainee perspective. Counselling Psychology Quarterly 2020 Jun 28:1-11. [doi: 10.1080/09515070.2020.1786668]
101. Schlegl S, Meule A, Favreau M, Voderholzer U. Bulimia nervosa in times of the COVID-19 pandemic-Results from an online survey of former inpatients. Eur Eat Disord Rev 2020 Nov 07;28(6):847-854 [FREE Full text] [doi: 10.1002/erv.2773] [Medline: 32767858]
102. Sciarrino NA, Myers US, Wangelin BC. When chaos is the norm: How some veterans with PTSD are continuing to engage in trauma-focused treatments during the COVID-19 pandemic. Psychol Trauma 2020 Aug;12(S1):S69-S70. [doi: 10.1037/tra0000718] [Medline: 32463289]
103. Sequeira A, Alozie A, Fasteau M, Lopez AK, Sy J, Turner KA, et al. Transitioning to virtual programming amidst COVID-19 outbreak. Counselling Psychology Quarterly 2020 Jun 12:1-16. [doi: 10.1080/09515070.2020.1777940]
104. Severe J, Tang R, Horbatch F, Onishchenko R, Naini V, Blazek MC. Factors Influencing Patients' Initial Decisions Regarding Telepsychiatry Participation During the COVID-19 Pandemic: Telephone-Based Survey. JMIR Form Res 2020 Dec 22;4(12):e25469 [FREE Full text] [doi: 10.2196/25469] [Medline: 33320823]
105. Sheehan R, Dalton-Locke C, Ali A, Vera San Juan N, Totsika V, Hassiotis A. Effects of the COVID-19 pandemic on mental healthcare and services: results of a UK survey of front-line staff working with people with intellectual disability and/or autism. BJPsych Bull 2021 May 12:1-7 [FREE Full text] [doi: 10.1192/bjb.2021.52] [Medline: 33977886]
106. Sklar M, Reeder K, Carandang K, Ehrhart M, Aarons G. An Observational Study of the Impact of COVID-19 and the Transition to Telehealth on Community Mental Health Center Providers. Res Sq 2020 Jul 29 [FREE Full text] [doi: 10.21203/rs.3.rs-48767/v1] [Medline: 32793893]
107. Termorshuizen JD, Watson HJ, Thornton LM, Borg S, Flatt RE, MacDermod CM, et al. Early impact of COVID-19 on individuals with self-reported eating disorders: A survey of ~1,000 individuals in the United States and the Netherlands. Int J Eat Disord 2020 Nov;53(11):1780-1790. [doi: 10.1002/eat.23353] [Medline: 32720399]
108. Uscher-Pines L, Sousa J, Raja P, Mehrotra A, Barnett ML, Huskamp HA. Suddenly Becoming a "Virtual Doctor": Experiences of Psychiatrists Transitioning to Telemedicine During the COVID-19 Pandemic. Psychiatr Serv 2020 Nov 01;71(11):1143-1150 [FREE Full text] [doi: 10.1176/appi.ps.202000250] [Medline: 32933411]
109. Uscher-Pines L, Sousa J, Raja P, Mehrotra A, Barnett M, Huskamp HA. Treatment of opioid use disorder during COVID-19: Experiences of clinicians transitioning to telemedicine. J Subst Abuse Treat 2020 Nov;118:108124 [FREE Full text] [doi: 10.1016/j.jsat.2020.108124] [Medline: 32893047]
110. van Dijk SDM, Bouman R, Folmer EH, den Held RC, Warringa JE, Marijnissen RM, et al. (Vi)-rushed Into Online Group Schema Therapy Based Day-Treatment for Older Adults by the COVID-19 Outbreak in the Netherlands. Am J Geriatr Psychiatry 2020 Sep;28(9):983-988 [FREE Full text] [doi: 10.1016/j.jagp.2020.05.028] [Medline: 32622730]
111. Wang X, Gordon RM, Snyder EW. Comparing Chinese and US practitioners' attitudes towards teletherapy during the COVID-19 pandemic. Asia Pac Psychiatry 2021 Mar 27;13(1):e12440 [FREE Full text] [doi: 10.1111/appy.12440] [Medline: 33108828]
112. Wilson C, Dalton-Locke C, Johnson S, Simpson A, Oram S, Howard L. Challenges and opportunities of the COVID-19 pandemic for perinatal mental health care: a mixed-methods study of mental health care staff. Arch Womens Ment Health 2021 Oct;24(5):749-757 [FREE Full text] [doi: 10.1007/s00737-021-01108-5] [Medline: 33830374]
113. Wood HJ, Gannon JM, Chengappa KR, Sarpal DK. Group teletherapy for first-episode psychosis: Piloting its integration with coordinated specialty care during the COVID-19 pandemic. Psychol Psychother 2021 Jun 20;94(2):382-389. [doi: 10.1111/papt.12310] [Medline: 33078496]

114. Wyler H, Liebrenz M, Ajdacic-Gross V, Seifritz E, Young S, Burger P, et al. Treatment provision for adults with ADHD during the COVID-19 pandemic: an exploratory study on patient and therapist experience with on-site sessions using face masks vs. telepsychiatric sessions. BMC Psychiatry 2021 May 05;21(1):237 [FREE Full text] [doi: 10.1186/s12888-021-03236-9] [Medline: 33952229]
115. Yellowlees P, Nakagawa K, Pakyurek M, Hanson A, Elder J, Kales HC. Rapid Conversion of an Outpatient Psychiatric Clinic to a 100% Virtual Telepsychiatry Clinic in Response to COVID-19. Psychiatr Serv 2020 Jul 01;71(7):749-752. [doi: 10.1176/appi.ps.202000230] [Medline: 32460683]
116. Zulfic Z, Liu D, Lloyd C, Rowan J, Schubert KO. Is telepsychiatry care a realistic option for community mental health services during the COVID-19 pandemic? Aust N Z J Psychiatry 2020 Dec 22;54(12):1228-1228 [FREE Full text] [doi: 10.1177/0004867420937788] [Medline: 32571079]
117. Turgoose D, Ashwick R, Murphy D. Systematic review of lessons learned from delivering tele-therapy to veterans with post-traumatic stress disorder. J Telemed Telecare 2018 Oct;24(9):575-585. [doi: 10.1177/1357633X17730443] [Medline: 28958211]
118. Norwood C, Moghaddam NG, Malins S, Sabin-Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: A systematic review and noninferiority meta-analysis. Clin Psychol Psychother 2018 Nov;25(6):797-808. [doi: 10.1002/cpp.2315] [Medline: 30014606]

Abbreviations

AACODS: authority, accuracy, coverage, objectivity, date, significance (tool)
CFIR: Consolidated Framework for Implementation Research
CMHT: community mental health team
MMAT: Mixed Methods Appraisal Tool
NHS: National Health Service
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by C Basch; submitted 05.07.21; peer-reviewed by S Bhattacharjya; comments to author 04.08.21; revised version received 20.09.21; accepted 01.10.21; published 09.12.21.

Please cite as:
Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, Sheridan Rains L, Goulding L, Badhan M, Roxburgh E, Barnett P, Spyridonidis S, Tomaskova M, Mo J, Harju-Seppänen J, Haime Z, Casetta C, Papamichail A, Lloyd-Evans B, Simpson A, Sevdalis N, Gaughran F, Johnson S
Implementation, Adoption, and Perceptions of Telemental Health During the COVID-19 Pandemic: Systematic Review
J Med Internet Res 2021;23(12):e31746
URL: https://www.jmir.org/2021/12/e31746
doi:10.2196/31746
PMID:34709179

Original Paper

Effects of Person-Centered Care Using a Digital Platform and Structured Telephone Support for People With Chronic Obstructive Pulmonary Disease and Chronic Heart Failure: Randomized Controlled Trial

Lilas Ali[1,2,3], PhD; Sara Wallström[1,2], PhD; Andreas Fors[1,2,4], PhD; Emmelie Barenfeld[1,2,5], PhD; Eva Fredholm[6]; Michael Fu[1,7], MD, PhD; Mahboubeh Goudarzi[2], MSc; Hanna Gyllensten[1,2], PhD; Irma Lindström Kjellberg[1,2], MSc; Karl Swedberg[1,2,7], MD, PhD; Lowie E G W Vanfleteren[8,9], MD, PhD; Inger Ekman[1,2], PhD

[1]Sahlgrenska Academy Institute of Health and Care Sciences, University of Gothenburg, Gothenburg, Sweden

[2]Centre for Person-Centred Care, University of Gothenburg, Gothenburg, Sweden

[3]Psychiatric Department, Sahlgrenska University Hospital, Gothenburg, Sweden

[4]Region Västra Götaland, Research and Development Primary Health Care, Gothenburg, Sweden

[5]Department of Occupational Therapy and Physiotherapy, Sahlgrenska University Hospital, Gothenburg, Sweden

[6]The Swedish Heart & Lung Foundation, Stockholm, Sweden

[7]Department of Molecular and Clinical Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[8]Chronic Obstructive Pulmonary Disease Center, Department of Respiratory Medicine and Allergology, Sahlgrenska University Hospital, Gothenburg, Sweden

[9]Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Corresponding Author:
Lilas Ali, PhD
Sahlgrenska Academy Institute of Health and Care Sciences
University of Gothenburg
Box 457
Gothenburg, 405 30
Sweden
Phone: 46 707587856
Email: lilas.ali@gu.se

Abstract

Background: Chronic obstructive pulmonary disease (COPD) and chronic heart failure (CHF) are characterized by severe symptom burden and common acute worsening episodes that often require hospitalization and affect prognosis. Although many studies have shown that person-centered care (PCC) increases self-efficacy in patients with chronic conditions, studies on patients with COPD and CHF treated in primary care and the effects of PCC on the risk of hospitalization in these patients are scarce.

Objective: The aim of this study is to evaluate the effects of PCC through a combined digital platform and telephone support for people with COPD and CHF.

Methods: A multicenter randomized trial was conducted from 2018 to 2020. A total of 222 patients were recruited from 9 primary care centers. Patients diagnosed with COPD, CHF, or both and with internet access were eligible. Participants were randomized into either usual care (112/222, 50.5%) or PCC combined with usual care (110/222, 49.5%). The intervention's main component was a personal health plan cocreated by the participants and assigned health care professionals. The health care professionals called the participants in the intervention group and encouraged narration to establish a partnership using PCC communication skills. A digital platform was used as a communication tool. The primary end point, divided into 2 categories (improved and deteriorated or unchanged), was a composite score of change in general self-efficacy and hospitalization or death 6 months after randomization. Data from the intention-to-treat group at 3- and 6-month follow-ups were analyzed. In addition, a per-protocol analysis was conducted on the participants who used the intervention.

Results: No significant differences were found in composite scores between the groups at the 3- and 6-month follow-ups. However, the per-protocol analysis of the 3-month follow-up revealed a significant difference in composite scores between the

study groups (*P*=.047), although it was not maintained until the end of the 6-month follow-up (*P*=.24). This effect was driven by a change in general self-efficacy from baseline.

Conclusions: PCC using a combined digital platform and structured telephone support seems to be an option to increase the short-term self-efficacy of people with COPD and CHF. This study adds to the knowledge of conceptual innovations in primary care to support patients with COPD and CHF.

Trial Registration: ClinicalTrials.gov NCT03183817; http://clinicaltrials.gov/ct2/show/NCT03183817

(J Med Internet Res 2021;23(12):e26794) doi:10.2196/26794

KEYWORDS

chronic heart failure; chronic obstructive pulmonary disease; digital platform; eHealth; patient-centered care; person-centered care; randomized controlled trial; telehealth

Introduction

Background

Chronic obstructive pulmonary disease (COPD) and chronic heart failure (CHF) are known for their high mortality and severe impact on daily living activities [1-3]. Although pharmacological therapy has dramatically improved outcomes over the past decade, patients still perceive a high symptom burden and acute worsening of events. Therefore, self-management strategies that enhance self-efficacy are crucial to optimize [4] and strengthen preventive approaches in primary care [5]. Digital solutions have been suggested as a safe option for addressing health care challenges and promoting self-management of chronic conditions such as COPD and CHF [6-8]. However, most digital solutions lack user involvement in the development of the platform [9].

Person-centered care (PCC) is an approach based on ethical principles by which a contractual agreement is formed involving the patient as an active partner in the care and decision-making process [10]. To support the operationalization of person-centered ethics in clinical practice, a framework was developed by the Gothenburg Centre for Person-Centred Care. This Gothenburg Centre for Person-Centered Care framework underlines the importance of cocreated care between patients and health care professionals (HCPs; eg, registered nurses and a physiotherapist) based on the patient's narrative, which identifies personal resources and potential barriers together with medical status [10,11]. A central concept of PCC is self-efficacy, that is, a person's conviction in his or her ability to manage challenges and complete a task successfully [12]. Enhanced self-efficacy has been shown to improve disease management and clinical outcomes, including health status in patients with chronic diseases [13], physical functioning in patients with COPD and CHF [14], and daily living in patients with COPD [15]. Thus, HCPs need to target patients' self-efficacy beliefs to perform desired activities and support them in taking responsibility for engaging in their care [16]. Previous research has shown that PCC increases the self-efficacy of patients [17-19]. PCC via telephone is also thought to mitigate worsening self-efficacy in COPD and CHF, indicating that a partnership could be established between patients and HCPs without face-to-face contact [20].

Several studies have used telemedicine and digital interventions for people with COPD and CHF [21,22]; however, PCC was not part of their design, and the results were mixed [23].

Objective

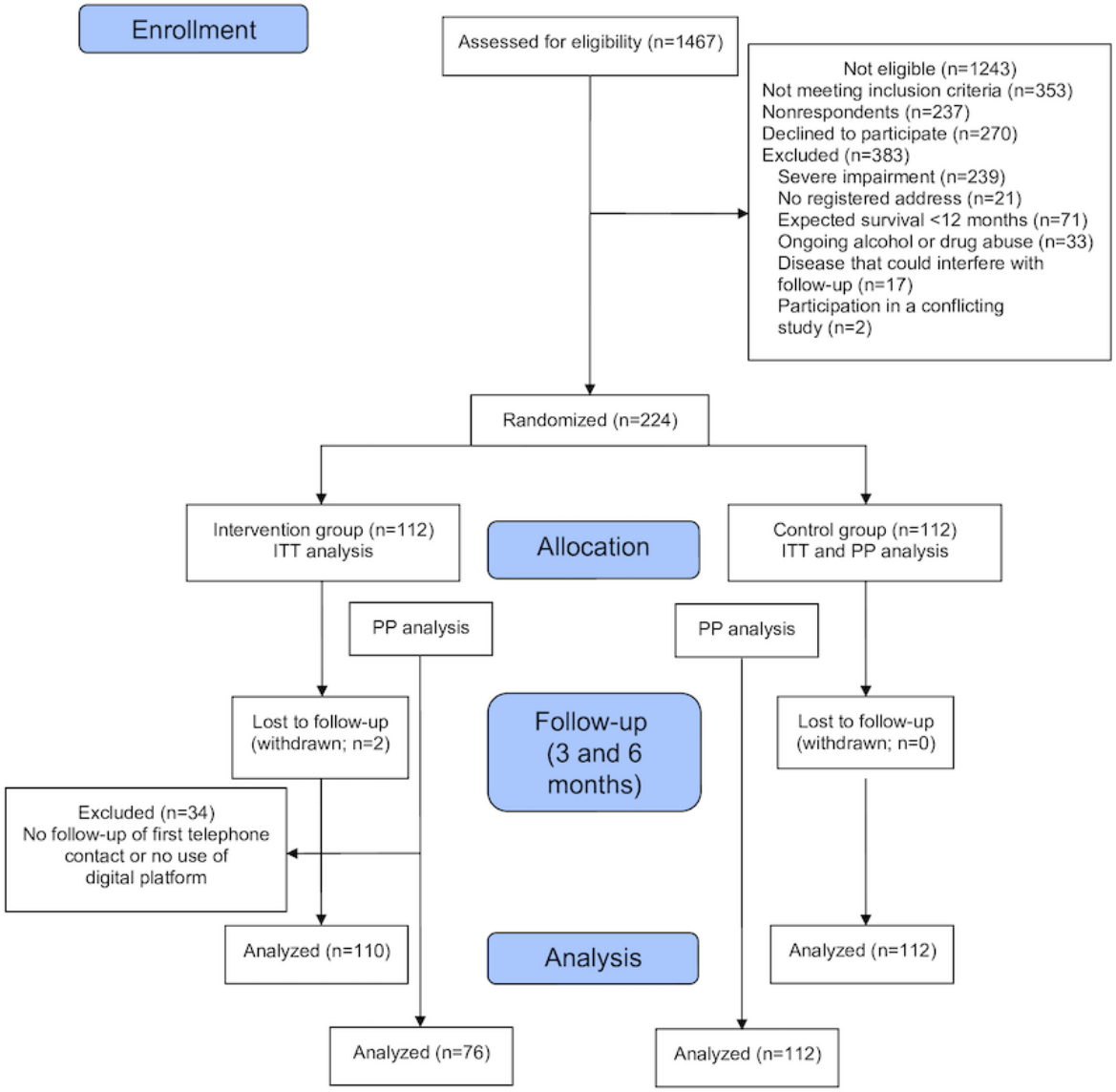
We hypothesize that PCC principles that include a digital platform and structured telephone support for people with COPD and CHF would reduce the need for primary care and hospital admission and improve self-efficacy through collaboration in the care process. Therefore, the aim of this study is to evaluate the effects of PCC through a combined digital platform and telephone support for people with COPD and CHF.

Methods

Design, Participants, and Setting

Study participants were recruited from 9 primary health care centers in Gothenburg from August 2017 to June 2019. The study, including consent forms for participants, was approved by the regional ethics board in Gothenburg (Dnr 063-17 and T613-18). The inclusion criteria were a diagnosis of COPD (J43.0, J44.0-J44.9) or CHF (I50.0-I50.9), being listed at one of the participating primary health care centers, understanding written and spoken Swedish, and having access to a device with an internet connection. The exclusion criteria were severe impairment (cognitive or other) that prevented the individual from using eHealth support, no registered address (follow-up questionnaires were sent by mail), expected survival of <12 months, ongoing documented diagnosis of alcohol or drug abuse, diseases that could interfere with follow-up (ie, multimorbidity), and participation in a conflicting study. A flowchart of the study participants is given in Figure 1.

Figure 1. CONSORT flowchart. ITT: intention-to-treat; PP: per-protocol.



Enrollment and Randomization

Designated HCPs screened medical records at the 9 participating centers for potential patients diagnosed with COPD, CHF, or both. Eligible participants were sent an information letter about the study with an invitation to contact the HCPs for more information. If the participants did not contact an HCP within 2 weeks after the information letter had been sent out, they were contacted by phone for further details and asked to participate. Patients who accepted the invitation to participate were sent a consent form together with a prepaid return envelope by mail. Upon receipt of the signed consent form, patients were randomized into either standard care or PCC in addition to standard care. Randomization was based on a computer-generated list created by a third party and stratified by age (<65 or ≥65 years) and diagnostic group (COPD, CHF, or COPD and CHF). All participants were informed of their allocation by phone, a call not included as part of the intervention.

Usual Care

Usual care was managed as per the physician's judgment based on current guidelines, for example, medicine adjustments [24,25]. The usual care group had no follow-up phone conversations.

Intervention

The intervention, comprising PCC using a combined digital platform and structured telephone support system, was provided in addition to usual care for 6 months. The structured telephone support program comprised an optional number of phone calls that included a health plan cocreated and followed up by patients and HCPs, which is consistent with person-centered principles. The digital platform was aimed at supporting communication between phone calls and providing access to shared documentation (health plans and self-ratings) and reliable information sources. The digital platform was developed using a participatory design with user involvement. The method draws on the user's *tacit knowledge*, that is, their implicit or unarticulated knowledge learned and transmitted through experience and apprenticeship, for example, by taking part in

this project. The researchers had workshops together with the HCPs, patient partners, and experts in which the platform and intervention were discussed and developed [26].

HCPs assisted participants in the intervention group in creating a log-in to the digital platform that they could access during the study period and described its features. In the first telephone conversations, the HCPs encouraged narration. They established a partnership using communication skills such as listening to the participants' narratives about daily life events and how they were affected by their condition. The next step entailed cocreating a health plan based on patient narratives through discussion and agreement, including patient goals, resources, and needs. Usually, the health plan contained information about what the participants had talked about, how they felt, what goals they had, and what they wanted to accomplish. The health plan also included information on the participants' capabilities and resources that could be used to help them reach their respective health goals. The health plan was written jointly at the initiative of the HCPs or the patients and uploaded to the digital platform by either the HCPs or the patients.

The HCPs and patients jointly scheduled the date of the follow-up meetings based on the preferences of the patients. The health plan served as a point of departure for the impending conversations and communication via the platform. The participants and HCPs had access to the platform during the 6-month study period. The health plan was considered and revised during each follow-up phone call and when needed (eg, if the participants spontaneously contacted the HCPs). The fidelity of the intervention was ensured by meetings and seminars on a regular basis, constructive discussions and education on PCC, and person-centered communication and monitoring of the HCPs by researchers and experts in their respective fields. The review group comprised senior and junior researchers and patient representatives. In addition, the HCPs reviewed some of each other's telephone calls and health plans. A total of 5 HCPs were involved in the intervention: 3 (60%) registered nurses, 1 (20%) occupational therapist, and 1 (20%) physiotherapist. Years of work experience ranged from 6 to 26 years for the HCPs.

The platform contained functionalities for 2-way communication through private messages; the possibility to rate daily symptoms, such as shortness of breath and tiredness, to be visualized as trend graphs; and an archive of the health plans. The participants could invite and give customized access to the platform to any person they wanted, such as informal carers, family, or friends. They could also access links to relevant websites containing information on COPD and CHF provided by patient organizations (eg, the Heart and Lung Association) and the Swedish national support guide to an online peer-to-peer support group. A detailed description of the intervention has been published elsewhere [27].

Collection of Data and Outcome Measures

Data were collected through questionnaires and from medical records at inclusion and 3 and 6 months later. Questionnaires were mailed to all participants together with a prepaid return envelope. If the questionnaires were not returned within 2 weeks, reminders were given by phone. New questionnaires and return envelopes were sent out if needed. Approximately 6.8% (15/222) of participants (11/110, 10% in the intervention group and 4/112, 3.6% in the control group) did not return their questionnaires at the 3-month follow-up. At the 6-month follow-up, 8.6% (19/222) of participants (13/110, 11.8% in the intervention group and 6/112, 5.4% in the control group) did not return their questionnaires despite being reminded.

The primary end point was a composite score of general self-efficacy (GSE) changes and hospitalization or death 6 months after randomization into each group.

A patient was classified as improved if GSE increased by ≥5 units, the patient was not hospitalized because of COPD or CHF, and the patient did not die. A patient was classified as deteriorated if GSE decreased by ≥5 units, the patient was admitted for unscheduled reasons because of COPD or CHF, or the patient died because of any cause.

Those who neither improved nor deteriorated were considered unchanged. GSE was assessed using the GSE scale, a 10-item scale designed to measure a sense of personal competence in dealing with stressful situations (eg, handling unforeseen situations and finding possible solutions to problems). The GSE scale has been widely tested and used internationally; the Swedish version has shown high internal consistency (Cronbach α=.90) [28]. The 10 items are rated on a 4-point scale, with total scores ranging from 10 to 40 [29]. An increase of 5 units has been suggested as a threshold for minimal clinically meaningful change [18,20].

Statistical Methods

A sample size of at least 91 participants in each group was needed to achieve a power of 80% based on a *P* value of .05 (2-tailed) to detect an increase in the composite score from 20% in the control group to 40% in the intervention group.

Descriptive and comparative statistics were used to characterize the study groups. Group differences were calculated using the Pearson chi-square test for categorical variables, the Fisher exact test for dichotomous variables, and the independent 2-tailed Student *t* test for continuous variables. Between-group differences in the composite score were tested using the Fisher exact test for the dichotomous version and the Mantel-Haenszel chi-square test for the ordered categorical version. Binary logistic regression was used to calculate odds ratios with 95% CIs for the dichotomous version of the composite score. The Student *t* test was used to compare the mean change in GSE scores between groups. Between-group differences in improvement of ≥5 points on the GSE scale were calculated in the same way as the dichotomous version of the composite score. Bivariate correlations were computed using Pearson *r*. Missing outcome data for the 3- and 6-month follow-ups were imputed using the last value carried forward. Sensitivity analyses were conducted to assess robustness. Both intention-to-treat (ITT) and per-protocol (PP) analyses were conducted. The PP group included participants with at least one PCC phone call and at least one health plan who logged into the platform and used at least one of its functions. The significance level was set at $P<.05$ (2-sided).

Results

Overview

In total, 224 participants were randomly assigned to either the control or the intervention group, of which 2 participants withdrew consent, leaving 222 participants (112/222, 50.5% in the control group and 110/222, 49.5% in the intervention group). The baseline characteristics are presented in Table 1. The study population was 53.6% (119/222) men and 46.4% (103/222) women, with a mean age of 70.8 (SD 9.4) years. Approximately 51.8% (115/222) participants had COPD, 38.3% (85/222) had CHF, and 9.9% (22/222) had both. Of the 222 participants, 32 (14.4%) were current smokers. The treatment and control groups were similar at baseline, except that significantly more participants in the control group were married or living with a partner than in the intervention group (P=.01). However, this difference did not remain when comparing the control and PP groups (P=.07). There were no significant differences between the groups in medical histories (eg, diagnosis, previous cardiovascular disease, or stages of COPD; Table 1).

Table 1. Baseline participant characteristics (N=222).

Characteristics	Control group (n=112)	ITT[a] group (n=110)		PP[b] group (n=76)	
		Value	*P* value	Value	*P* value
Age (years), mean (SD)	70.4 (9.1)	71.1 (9.8)	.59	70.1 (9.1)	.84
Women, n (%)	52 (46.4)	51 (46.4)	.99	31 (40.8)	.46
BMI, mean (SD)	28.2 (5.1)[c]	29.0 (5.4)[d]	.33	29.0 (5.3)[e]	.35
GSE[f] score, mean (SD)	31.1 (6.2)	31.0 (5.4)[g]	.99	30.9 (5.4)[h]	.83
Civil status, n (%)					
Living alone	25 (22.3)	42 (38.2)	.01	25 (34.2)	.07
Married or living with partner	87 (77.7)	68 (61.8)	.01	49 (64.5)	.07
Diagnosis, n (%)					
CHF[i]	43 (38.4)	42 (38.2)	.88	27 (35.5)	.49
COPD[j]	59 (52.7)	56 (50.9)	.88	38 (50)	.49
CHD and COPD	10 (8.9)	12 (10.9)	.88	11 (14.5)	.49
Stage of COPD, n (%)					
Stage 1	16 (27.6)	16 (26.2)	.99	10 (22.2)	.94
Stage 2	36 (62.1)	38 (62.3)	.99	30 (66.7)	.94
Stage 3	5 (8.6)	6 (9.8)	.99	4 (8.9)	.94
Stage 4	1 (1.7)	1 (1.6)	.99	1 (2.2)	.94
Education level, n (%)					
Compulsory	26 (23.2)	38 (34.5)	.13	26 (34.2)	.23
Secondary school	32 (28.6)	25 (22.7)	.13	18 (23.7)	.23
Vocational college	21 (18.8)	25 (22.7)	.13	17 (22.4)	.23
University	33 (29.5)	22 (20)	.13	15 (19.7)	.23
Smoking, n (%)					
Current smoker	19 (17)	13 (11.8)	.51	8 (10.5)	.47
Previous smoker	63 (56.3)	63 (57.3)	.51	46 (60.5)	.47
Never smoked	30 (26.8)	34 (30.9)	.51	22 (28.9)	.47
Medical history, n (%)					
Previous MI[k]	12 (10.7)	15 (13.6)	.54	10 (13.2)	.65
Previous angina	8 (7.1)	9 (8.2)	.81	8 (10.5)	.44
Atrial fibrillation	31 (27.7)	39 (35.5)	.25	26 (34.2)	.42
Hypertension	68 (60.7)	76 (69.1)	.21	49 (64.5)	.65
CABG[l]	3 (2.7)	2 (1.8)	.99	2 (2.6)	.99
Previous stroke	6 (5.4)	11 (10)	.22	10 (13.2)	.07
Diabetes	19 (17)	23 (20.9)	.50	18 (23.7)	.27
CRT[m]	1 (0.9)	1 (0.9)	.99	1 (1.3)	.99
Pacemaker	7 (6.3)	6 (5.5)	.99	4 (5.3)	.99
Previous spirometry <6 months before inclusion, n (%)	19 (31.7)	14 (22.6)	.31	11 (25)	.52
<6 months FEV[n] 1% of expected value, mean (SD)	68.8 (17.3)[o]	67.6 (16.9)[p]	.85	64.1 (16.0)[q]	.47
Previous spirometry 6-12 months before inclusion, n (%)	8 (13.3)	13 (21)	.34	11 (25)	.20

Characteristics	Control group (n=112)	ITT[a] group (n=110)		PP[b] group (n=76)	
		Value	*P* value	Value	*P* value
6-12 months FEV 1% of expected value, mean (SD)	66.8 (14.0)[r]	70.5 (15.6)[s]	.61	68.7 (16.4)[t]	.81
Previous spirometry >12 months before inclusion, n (%)	31 (54.4)	33 (55)	.99	22 (52.4)	.99
>12 months FEV 1% of expected value, mean (SD)	71.0 (16.4)[u]	66.1 (16.2)[v]	.24	64.1 (16.2)[w]	.14

[a]ITT: intention-to-treat.
[b]PP: per-protocol.
[c]n=91.
[d]n=87.
[e]n=60.
[f]GSE: general self-efficacy.
[g]n=107.
[h]n=73.
[i]CHF: chronic heart failure.
[j]COPD: chronic obstructive pulmonary disease.
[k]MI: myocardial infarction.
[l]CABG: coronary artery bypass graft.
[m]CRT: cardiac resynchronization therapy.
[n]FEV: forced expiratory volume.
[o]n=19.
[p]n=13.
[q]n=11.
[r]n=7.
[s]n=12.
[t]n=10.
[u]n=30.
[v]n=33.
[w]n=21.

During the 6-month intervention, the ITT group had a median of 4 (range 0-11) telephone conversations with the HCPs. Of those, a median of 3 (range 0-7) was within 90 days of randomization. The PP group had a median of 4 (range 1-11) telephone conversations during the study period, with a median of 3 (range 1-7) within 90 days of randomization. No statistically significant correlations were detected between the number of calls and changes in GSE.

Effects

No significant differences in composite scores (improved vs deteriorated or unchanged) at the 3- or 6-month follow-ups were observed between the groups. However, a significant difference in the PP analysis was noted at the 3-month follow-up (*P*=.047). The analysis confirmed this result using the composite score (improved, unchanged, or deteriorated; *P*=.04). However, none of these differences could be sustained at the 6-month follow-up (*P*=.24 and *P*=.50; Table 2). We found no differences in composite scores between patients with COPD, CHF, or COPD and CHF. During the 6-month intervention, there were 4 recorded hospitalizations: 3 (75%) in the intervention group and 1 (25%) in the control group. There were no participant deaths during the intervention.

Table 2. Composite scores at the 3- and 6-month follow-up.

Time and composite score	Control (n=112)	Intention-to-treat (n=110)[a]			Per-protocol (n=76)[a]		
		Value	OR[b] (95% CI)	*P* value	Value	OR (95% CI)	*P* value
3 months							
Composite score 1[c], n (%)							
Improved	10 (8.9)	15 (14)	1.663 (0.712-3.884)	.24	14 (19.2)	2.420 (1.011-5.792)	.047[d]
Deteriorated or unchanged	102 (91.1)	92 (86)	1.663 (0.712-3.884)	.24	59 (80.8)	2.420 (1.011-5.792)	.047[d]
Composite score 2[e], n (%)			—[f]	—		—	
Improved	10 (8.9)	15 (14)			14 (19.2)		.04[d]
Unchanged	89 (79.5)	83 (77.6)			54 (74)		.04[d]
Deteriorated	13 (11.6)	9 (8.4)			5 (6.8)		.04[d]
6 months							
Composite score 1[c], n (%)							
Improved	13 (11.6)	16 (15)	1.339 (0.611-2.936)	.47	13 (17.8)	1.650 (0.717-3.795)	.24
Deteriorated or unchanged	99 (88.4)	91 (85)	1.339 (0.611-2.936)	.47	60 (82.2)	1.650 (0.717-3.795)	.24
Composite score 2[e], n (%)			—	—		—	
Improved	13 (11.6)	16 (15)			13 (17.8)		.50
Unchanged	83 (74.1)	74 (69.2)			49 (67.1)		.50
Deteriorated	16 (14.3)	17 (15.9)			11 (15.1)		.50

[a]3 missing values (no general self-efficacy score at baseline).

[b]OR: odds ratio.

[c]Composite score dichotomized into improved vs deteriorated or unchanged.

[d]Significant at $P<.05$.

[e]Composite score improved vs unchanged vs deteriorated.

[f]The statistical test used (Mantel-Haenszel chi-square test) does not generate odds ratio or 95% CI.

Differences between the groups at 3 months were driven mainly by changes in the GSE scale. The PP group had a mean improvement of 0.941 (SD 4.4) points, whereas the control group had a mean reduction of 0.568 points (SD 4.6; $P=.03$). At the 3-month follow-up, 19.2% (14/76) of the participants in the PP group versus 8.9% (10/112) of those in the control group had an improvement of ≥5 points on the GSE scale, indicating that the PP group was more than twice as likely to have a clinically meaningful improvement in GSE compared with the control group (Table 3).

Table 3. Change in general self-efficacy (GSE) from baseline and at the 3- and 6-month follow-ups.

Follow-up	Control (n=112)	ITT[a] (n=110)[b]			PP[c] (n=76)[d]		
		Value	OR[e] (95% CI)	*P* value	Value	OR (95% CI)	*P* value
3 months							
Change in GSE score, mean (SD)	−0.568 (4.6)	0.544 (4.0)	2.261 to 0.038[d]	.06	0.941 (4.4)	−2.842 to −0.174[d]	.03[f]
Improvement ≥5 points, n (%)	10 (8.9)	15 (14)	1.663 (0.712 to 3.884)	.24	14 (19.2)	2.420 (1.011 to 5.792)	.047[f]
6 months							
Change in GSE score, mean (SD)	−0.397 (4.5)	0.127 (4.3)	−1.689 to 0.642[d]	.38	0.384 (4.4)	−2.103 to 0.542[d]	.25
Improvement ≥5 points, n (%)	13 (11.6)	16 (15)	1.339 (0.611 to 2.936)	.47	13 (17.8)	1.650 (0.717 to 3.795)	.24

[a]ITT: intention-to-treat.

[b]3 missing values (no general self-efficacy score at baseline).

[c]PP: per-protocol.

[d]Only 95% CI shown as the statistical test (Student *t* test) does not generate odds ratio.

[e]OR: odds ratio.

[f]Significant at $P<.05$.

Discussion

Principal Findings

This study shows that a 6-month intervention with PCC using a combined digital platform and structured telephone support system does not improve outcome as assessed by a composite end point of change on the GSE score and absence of hospitalization or death. However, the PP analysis, which only included those who used the intervention, showed a difference in composite scores at 3 months. However, this difference did not hold at the 6-month follow-up. The difference at 3 months was primarily because of an improvement in the GSE scale score. PP analyses are usually data-driven and thus, of lower evidence than prespecified analyses but are important to explore the mechanisms behind the effects of treatment and care interventions [30].

Several explanations have been proposed to account for why the ITT analysis showed no significant differences and why the PP analysis only showed differences at the 3-month follow-up. The participants had a relatively stable disease, as illustrated by the few disease-related hospitalizations and the high self-reported GSE at baseline (mean score 31.1; Table 1). In contrast, population studies have reported a mean GSE score of 29 [29]. A previous study from our group targeting patients with more severe forms of COPD and CHF demonstrated more pronounced positive results from a similar intervention. In that study, the participants had lower GSE (mean score of 28) at baseline and therefore, had more margin for improvement [20]. Previous studies have highlighted the difficulties in showing the significance of an intervention when the participants' score on outcome measurements at baseline (ie, ceiling effects) was high [31].

The intervention's timing might have played a role in the lack of significant differences between the groups. The intervention sought to reach out with a preventive tool to support self-management and, therefore, recruited participants from primary care centers. Thus, unlike our previous interventions that mainly recruited participants during an unplanned hospitalization, in this intervention, patients were targeted in their everyday lives. In those interventions, hospitalization was an exact starting point, and many were eager to recover their health after a period of deterioration [18,20]. That the intervention's timing is important for its meaningfulness was also confirmed by the findings from the participant interviews [32]. Moreover, previous research has pointed out the importance of including participants in need of an intervention to achieve favorable outcomes [31]. In this study, screening was not performed to determine which participants might be most suitable for the intervention.

The intervention showed an effect after 3 months in the PP analysis but not after 6 months. This result is likely because of the participants' initial high degree of communication with the HCPs and the fact that the increase in GSE caused by the intervention attenuated over time. At least two phone meetings were initially made, one occurring when the patients were contacted and asked to participate and the other during the health plan's joint formulation. As the median number of calls was 4, many of the participants had few conversations with HCPs after the first 2 calls. The intervention did not include mandatory follow-up or booster calls. If this had been done, the effect of the intervention might have been maintained. Nevertheless, no statistically significant associations were found between the number of calls and changes in the GSE. Associations might have been present; however, the study was underpowered to detect them. There may also be other explanations for the lack of effect after 6 months, such as lack of motivation and

worsening of COPD and CHF; however, these variables were not collected.

Another possibility for the ITT analysis not detecting any differences is that the participants did not use the digital platform to its fullest extent. A reason for this lack of difference may be that the participants did not feel the need to use the digital platform when feeling stable [32]. Even if used, the digital platform might not have added enough support to improve GSE. These possibilities are important to explore further to find the best technique for future digital communication between HCPs and participants. It can be speculated that this essential partnership may not be established without face-to-face meetings. However, a process evaluation of the trial using grounded theory has shown that it is possible to establish partnerships without face-to-face meetings [32]. Although the intervention might not suit all people with COPD and CHF, those who used the platform and structured telephone support (ie, the PP group) showed a significant improvement in GSE. In addition, an increase of ≥5 on the GSE scale, which was considered a clinically meaningful improvement, was equivalent to almost 1 SD in our sample. In general, 0.5 SD is viewed as the cutoff for a clinically significant difference [33]. Previous research has shown that GSE improvement is associated with better food choices [34], improved exercise endurance in people with COPD [35], and functional fitness among older adults [36]. These findings indicate that GSE increment may support people with chronic illnesses to improve and maintain health. This study shows that PCC using a combined digital platform and structured telephone support system is one way to support people with COPD and CHF. However, this type of intervention must target those who would benefit the most from it. An important finding for HCPs is that when using self-care tools for home monitoring, our patients with CHF emphasized that they did not want to read or be reminded of CHF. This may be a way for people with CHF to cope with everyday life [8]. It also highlights the need for HCPs to listen to their patients' concerns and individualize care accordingly.

To our knowledge, this is the first randomized trial in which the effects of PCC using a combined digital platform and structured telephone support system for people with COPD and CHF treated in primary care were evaluated. This study adds knowledge on which factors (eg, timing of the intervention and sustainability of interventional effects) are essential when designing interventions targeting self-care. In general, most self-care interventions are time intensive and require effort from the interventionist. Information on the sustainability of the effects after the end of the intervention is often not reported. Nonetheless, telehealth self-management interventions do not seem to report any adverse effects, which could suggest that such interventions are a suitable option to support people with chronic conditions [6].

Strengths and Limitations

This study has several limitations worth noting. First, there is no information on the disease severity of CHF in the background information of the participants, as this was not consequently reported in medical records. Second, few hospitalization events indicate that another outcome measure might have been more suitable for this study population. Furthermore, the participants had a high mean GSE score at baseline, leaving little margin for improvement. The results might have been different if another outcome measure had been evaluated. Fourth, the design of the intervention might not have been ideal for all of the included participants. More participants might have been identified and benefited from the intervention if the study had used more rigorous screening criteria. Fifth, this manuscript only focuses on the effects of the intervention and might have been strengthened by adding a process evaluation; however, those data are, unfortunately, not currently available.

This study also has some strengths. First, the intervention improved GSE at the 3-month follow-up for the participants who used it. In addition, the use of telehealth made it more accessible and reduced the need for patients to travel to and from health care centers.

Conclusions

By combining a digital platform with structured telephone support, PCC seems to be an option to increase the short-term self-efficacy of people with COPD and CHF. This study adds to the knowledge of conceptual innovations in the primary care setting to support patients with COPD and CHF. Further research is needed to explore which patient at what point in the natural history of the disease would benefit the most and tailor different digital interventions and PCC components to each patient's unique needs.

Acknowledgments

The authors want to thank all the participants for sharing their experiences and the professional health care workers Eva-Lena Andersson, Paula Johansson Tunås, Emma Vasell, and Dr Joanne Fuller. The authors also gratefully acknowledge the contribution of the participating primary care centers, their managers, and particularly area manager Cathrine Thanner for their support and assistance in conducting the study.

This work was supported by The Swedish Heart and Lung Foundation (DNr.20180183), the Swedish Research Council (Dnr 2017-01230), and the Centre for Person-Centred Care at the University of Gothenburg, Sweden. The funders had no role in the analysis or preparation of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 323 KB - jmir_v23i12e26794_app1.pdf]

References

1. Theander K, Hasselgren M, Luhr K, Eckerblad J, Unosson M, Karlsson I. Symptoms and impact of symptoms on function and health in patients with chronic obstructive pulmonary disease and chronic heart failure in primary health care. Int J Chron Obstruct Pulmon Dis 2014 Jun 18;9:785-794 [FREE Full text] [doi: 10.2147/COPD.S62563] [Medline: 25071370]
2. Rabe KF, Watz H. Chronic obstructive pulmonary disease. Lancet 2017 May 13;389(10082):1931-1940. [doi: 10.1016/S0140-6736(17)31222-9] [Medline: 28513453]
3. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, ESC Committee for Practice Guidelines. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur Heart J 2012 Jul;33(14):1787-1847. [doi: 10.1093/eurheartj/ehs104] [Medline: 22611136]
4. Gold P. The 2007 GOLD guidelines: a comprehensive care framework. Respir Care 2009 Aug;54(8):1040-1049 [FREE Full text] [Medline: 19650945]
5. Zakrisson A, Arne M, Hasselgren M, Lisspers K, Ställberg B, Theander K. A complex intervention of self-management for patients with COPD or CHF in primary care improved performance and satisfaction with regard to own selected activities; A longitudinal follow-up. J Adv Nurs 2019 Jan;75(1):175-186. [doi: 10.1111/jan.13899] [Medline: 30375028]
6. 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]
7. World Health Organization. Atlas of EHealth Country Profiles: The Use of EHealth in Support of Universal Health Coverage: Based on the Findings of the Third Global Survery on EHealth 2015. Geneva: World Health Organization; 2016.
8. Aamodt IT, Strömberg A, Hellesø R, Jaarsma T, Lie I. Tools to support self-care monitoring at home: perspectives of patients with heart failure. Int J Environ Res Public Health 2020 Nov 30;17(23):8916 [FREE Full text] [doi: 10.3390/ijerph17238916] [Medline: 33266245]
9. Bringsvor HB, Langeland E, Oftedal BF, Skaug K, Assmus J, Bentsen SB. Effects of a COPD self-management support intervention: a randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2018;13:3677-3688 [FREE Full text] [doi: 10.2147/COPD.S181005] [Medline: 30510410]
10. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
11. Britten N, Ekman I, Naldemirci Ö, Javinger M, Hedman H, Wolf A. Learning from Gothenburg model of person centred healthcare. BMJ 2020 Sep 01;370:m2738. [doi: 10.1136/bmj.m2738] [Medline: 32873594]
12. 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]
13. 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]
14. Arnold R, Ranchor AV, DeJongste MJ, Köeter GH, Ten Hacken NH, Aalbers R, et al. The relationship between self-efficacy and self-reported physical functioning in chronic obstructive pulmonary disease and chronic heart failure. Behav Med 2005;31(3):107-115. [doi: 10.3200/BMED.31.3.107-115] [Medline: 16252622]
15. Kar S, Zengin N. The relation between self-efficacy in patients with chronic obstructive pulmonary disease and caregiver burden. Scand J Caring Sci 2020 Sep;34(3):754-761. [doi: 10.1111/scs.12780] [Medline: 31657060]
16. Lau-Walker M, Thompson DR. Self-management in long-term health conditions—a complex concept poorly understood and applied? Patient Educ Couns 2009 May;75(2):290-292. [doi: 10.1016/j.pec.2009.02.001] [Medline: 19272746]
17. Fors A, Taft C, Ulin K, Ekman I. Person-centred care improves self-efficacy to control symptoms after acute coronary syndrome: a randomized controlled trial. Eur J Cardiovasc Nurs 2016 Apr;15(2):186-194. [doi: 10.1177/1474515115623437] [Medline: 26701344]
18. Fors A, Ekman I, Taft C, Björkelund C, Frid K, Larsson ME, et al. Person-centred care after acute coronary syndrome, from hospital to primary care - A randomised controlled trial. Int J Cardiol 2015;187:693-699. [doi: 10.1016/j.ijcard.2015.03.336] [Medline: 25919754]
19. Ali L, Wallström S, Ekman I, Swedberg K, Fors A. Effects of person-centred care via telephone on self-efficacy in patients with chronic obstructive pulmonary disease: subgroup analysis of a randomized controlled trial. Nurs Open 2021 Mar;8(2):927-935 [FREE Full text] [doi: 10.1002/nop2.701] [Medline: 33570304]
20. Fors A, Blanck E, Ali L, Ekberg-Jansson A, Fu M, Kjellberg IL, et al. Effects of a person-centred telephone-support in patients with chronic obstructive pulmonary disease and/or chronic heart failure - a randomized controlled trial. PLoS One 2018 Aug 31;13(8):e0203031 [FREE Full text] [doi: 10.1371/journal.pone.0203031] [Medline: 30169539]

21. Barbosa MT, Sousa CS, Morais-Almeida M, Simões MJ, Mendes P. Telemedicine in COPD: an overview by topics. COPD 2020 Oct;17(5):601-617. [doi: 10.1080/15412555.2020.1815182] [Medline: 32892650]
22. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
23. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
24. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Rev Esp Cardiol (Engl Ed) 2016 Dec;69(12):1167. [doi: 10.1016/j.rec.2016.11.005] [Medline: 27894487]
25. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. Am J Respir Crit Care Med 2017 Mar 01;195(5):557-582. [doi: 10.1164/rccm.201701-0218PP] [Medline: 28128970]
26. Spinuzzi C. The methodology of participatory design. Technical Commun 2005;52(2):163-174.
27. Ali L, Wallström S, Barenfeld E, Fors A, Fredholm E, Gyllensten H, et al. Person-centred care by a combined digital platform and structured telephone support for people with chronic obstructive pulmonary disease and/or chronic heart failure: study protocol for the PROTECT randomised controlled trial. BMJ Open 2020 Jul 19;10(7):e036356 [FREE Full text] [doi: 10.1136/bmjopen-2019-036356] [Medline: 32690519]
28. Löve J, Moore CD, Hensing G. Validation of the Swedish translation of the General Self-Efficacy scale. Qual Life Res 2012 Sep;21(7):1249-1253. [doi: 10.1007/s11136-011-0030-5] [Medline: 21984467]
29. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Measures in Health Psychology: a User's Portfolio. Causal and Control Beliefs. Windsor, UK: NFER-NELSON; 1995.
30. Ekman I, Wolf A, Olsson L, Taft C, Dudas K, Schaufelberger M, et al. Effects of person-centred care in patients with chronic heart failure: the PCC-HF study. Eur Heart J 2012 May 15;33(9):1112-1119 [FREE Full text] [doi: 10.1093/eurheartj/ehr306] [Medline: 21926072]
31. Jaarsma T, Strömberg A, Dunbar SB, Fitzsimons D, Lee C, Middleton S, et al. Self-care research: how to grow the evidence base? Int J Nurs Stud 2020 May;105:103555 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103555] [Medline: 32199150]
32. Barenfeld E, Ali L, Wallström S, Fors A, Ekman I. Becoming more of an insider: a grounded theory study on patients' experience of a person-centred e-health intervention. PLoS One 2020 Nov 23;15(11):e0241801 [FREE Full text] [doi: 10.1371/journal.pone.0241801] [Medline: 33226986]
33. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003 May;41(5):582-592. [doi: 10.1097/01.MLR.0000062554.74615.4C] [Medline: 12719681]
34. Sol BG, van der Graaf Y, van Petersen R, Visseren FL. The effect of self-efficacy on cardiovascular lifestyle. Eur J Cardiovasc Nurs 2011 Sep;10(3):180-186. [doi: 10.1016/j.ejcnurse.2010.06.005] [Medline: 20674503]
35. Toshima M, Kaplan R, Ries A. Self-efficacy expectations in chronic obstructive pulmonary disease rehabilitation. In: Self-Efficacy: Thought Control of Action. New York: Hemisphere; 1992.
36. Konopack JF, Marquez DX, Hu L, Elavsky S, McAuley E, Kramer AF. Correlates of functional fitness in older adults. Int J Behav Med 2008;15(4):311-318 [FREE Full text] [doi: 10.1080/10705500802365557] [Medline: 19005931]

Abbreviations

CHF: chronic heart failure
COPD: chronic obstructive pulmonary disease
GSE: general self-efficacy
HCP: health care professional
ITT: intention-to-treat
PCC: person-centered care
PP: per-protocol

Edited by R Kukafka, G Eysenbach; submitted 15.01.21; peer-reviewed by W Stut, L Sheeran; comments to author 14.03.21; revised version received 15.04.21; accepted 13.10.21; published 13.12.21.

Please cite as:
Ali L, Wallström S, Fors A, Barenfeld E, Fredholm E, Fu M, Goudarzi M, Gyllensten H, Lindström Kjellberg I, Swedberg K, Vanfleteren LEGW, Ekman I
Effects of Person-Centered Care Using a Digital Platform and Structured Telephone Support for People With Chronic Obstructive Pulmonary Disease and Chronic Heart Failure: Randomized Controlled Trial
J Med Internet Res 2021;23(12):e26794
URL: https://www.jmir.org/2021/12/e26794
doi:10.2196/26794
PMID:34898447

Original Paper

One-Year Remission Rate of Chronic Headache Comparing Video and Face-to-Face Consultations by Neurologist: Randomized Controlled Trial

Svein Ivar Bekkelund[1], MD, PhD; Kai Ivar Müller[1], MD, PhD

Department of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Svein Ivar Bekkelund, MD, PhD
Department of Clinical Medicine
The Arctic University of Norway
Hansine Hansens veg 18
Tromsø, 9037
Norway
Phone: 47 90185090
Email: svein-ivar.bekkelund@uit.no

Abstract

Background: Chronic headache causing severe headache-related disability for those affected by the disease is under- or misdiagnosed in many cases and therefore requires easy access to a specialist for optimal health care management.

Objective: The goal of the research is to determine whether video consultations are noninferior to face-to-face consultations in treating chronic headache patients referred to a specialist in Northern Norway.

Methods: Patients included in the study were recruited from general practice referrals to a specialist at a neurological department in Northern Norway (Tromsø) and diagnosed according to the International Headache Society classification system. In a randomized controlled design, the 1-year remission rate of chronic headache (change from ≥15 to <15 headache days per month during the last 3 months), patient satisfaction with a specialist consultation, and need for follow-up consultations by general practitioners were compared between groups consulted by video and face-to-face in a post hoc analysis. Data were collected by interview (baseline) and questionnaire (follow-up).

Results: From a baseline cohort of 402 headache patients consecutively referred from general practice to a specialist over 2.5 years, 58.0% (233/402) were classified as chronic headache and included in this study. Response rates were 71.7% (86/120) in the video group and 67.3% (76/113) in the face-to-face group. One-year remission from chronic headache was achieved in 43.0% (37/86) in the video group and 39.5% (30/76) in the face-to-face group (P=.38). Patient satisfaction with consultations were 86.5% (32/37; video) and 93.3% (28/30; face-to-face; P=.25). A total of 30% (11/37) in the video group and 53% (16/30) in the face-to-face group consulted general practitioners during the follow-up period (P=.03), and median number of consultations was 1 (IQR 0-13) and 1.5 (IQR 0-15), respectively (P=.19).

Conclusions: One-year remission rate from chronic headache was about 40% regardless of consultation form. Likewise, patient satisfaction with consultation and need for follow-up visits in general practice post consultation was similar. Treating chronic headache patients by using video consultations is not inferior to face-to-face consultations and may be used in clinical neurological practice.

Trial Registration: ClinicalTrials.gov NCT02270177; https://clinicaltrials.gov/ct2/show/NCT02270177

(J Med Internet Res 2021;23(12):e30151) doi:10.2196/30151

KEYWORDS

chronic headache; remission; video consultation; telemedicine; eHealth; digital consultation; consultation; treatment; follow-up; RCT; randomized controlled trial

Introduction

Background

Chronic headache is a condition that transforms from primary headaches and is mainly identified as chronic migraine and chronic tension-type headache affecting about 1% to 2% and 2% of the population, respectively [1-3]. Lack of diagnostic biomarkers is a major challenge, and the chronic headache diagnosis (presence of headache 15 or more days per month for the last 3 months) is made by using a structured interview that relies on a validated diagnostic classification system [4]. Headache burden in the population is high and has not improved at the population level over time [5,6]. Preferably, headache should be correctly classified and treated at earliest to minimize risk of becoming chronic and development of associated conditions such as psychiatric symptoms [7], comorbid pain [8], increased costs for patients and the society [9], and impaired work performance [10], which are some known consequences that may burden chronic headache patients further. Depression, anxiety, sleep problems, stress, medication overuse, and low degree of headache self-management were associated with poorer prognosis of chronic headache in randomized controlled trials (RCTs) and prospective cohorts as summarized in an American review [11]. It has long been known that headache syndromes are underdiagnosed and undertreated, especially migraine [12,13]. There are many possible reasons for that. Professional headache care needs to be better coordinated with general health practice [14,15] and is further challenged by variable access to headache specialists [16]. In a group of 1254 chronic migraine sufferers, 40% had consulted a headache specialist, but only 4.5% reported that they additionally received correct diagnosis and treatment [13]. Knowledge about how alternative specialist consultations using information and communication technology may be used to treat patients with difficult headache is limited. In less populated areas where a secondary health service such as a general neurological department is the only alternative to primary health care, telemedicine might be effective. The main hypothesis of this study was that treating chronic headache patients referred to a neurological department by using video consultations is not inferior to traditional face-to-face consultations.

Methods

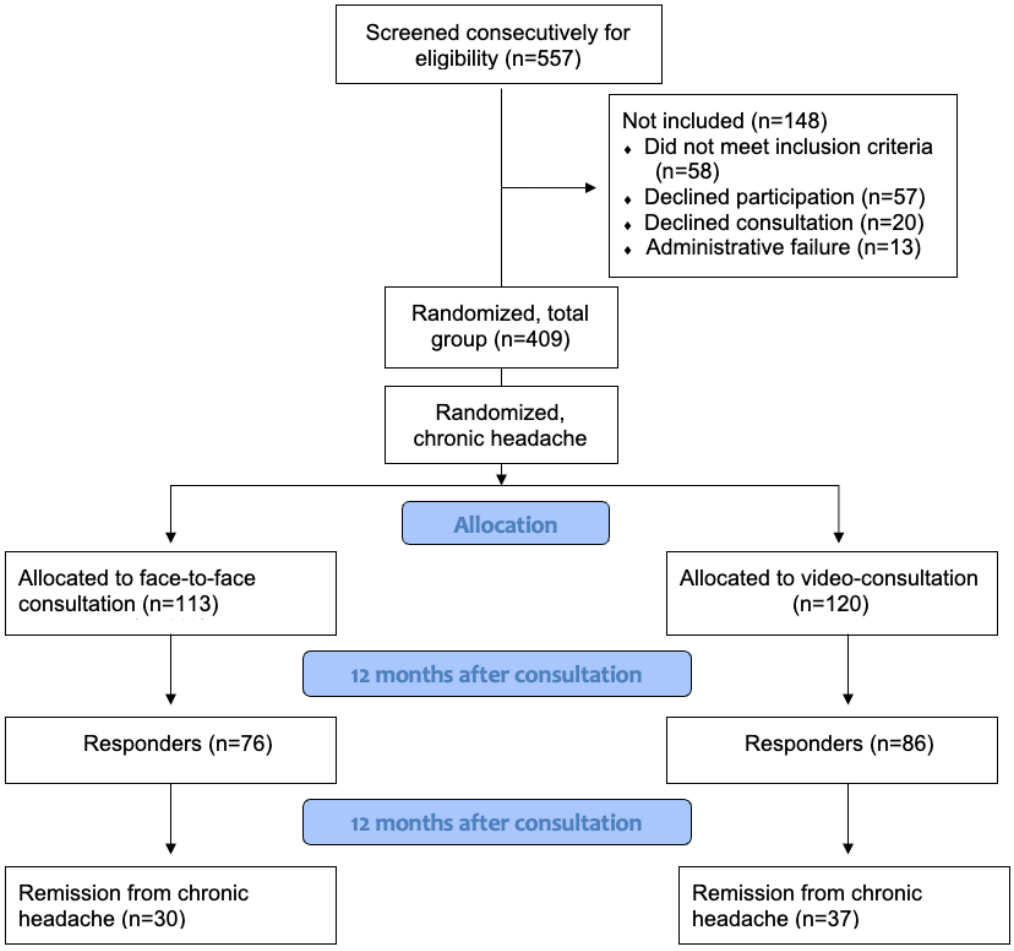
Study Design and Hypotheses

The results from this study are based on post hoc analyses from a previous open-label randomized clinical trial where a larger group (n=409) of heterogeneous headache patients referred from general practitioners (GPs) to specialists were assigned to either video or face-to-face (in-office) consultations to study cost, feasibility, and clinical aspects [17]. The trial was registered at ClinicalTrials.gov [NCT02270177]. The group of patients with chronic headache from that cohort were selected for this study to test the following primary hypothesis: providing treatment to chronic headache patients by specialist video consultations is noninferior to face-to-face specialist consultations with respect to 1-year remission rate of chronic headache (change from ≥15 to <15 headache days per month during the last 3 months). Secondary hypotheses were (1) patient satisfaction with specialist consultation, (2) frequency of chronic headache patients visiting GP for headache in the 12 months postconsultation, and (3) median number of headache-related consultations at GPs in the 12 months posttreatment, all end points postulated to be indifferent between the groups. The CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth) was used as a guide in describing the scientific study method [18].

Study Population and Randomization

Patients were consecutively identified, screened, randomized, and consulted for 2.5 years (September 30, 2012, to March 30, 2015). Of the included patients, 58.0% (233/402) of patients were classified to have chronic headache and included in the study (Figure 1).

Figure 1. Flowchart of patients with chronic headache referred to neurologists from general practitioners for headache fulfilling the study inclusion criteria.



Selection criteria were as follows: (1) patients referred from GP to neurologist for headache, (2) fulfilling the classification criteria for chronic headache without evidence of secondary headache (ie, headaches classified as primary headache without specific causes [4] except patients with suspected medication overuse headache), (3) Norwegian speaking men and women aged 16 to 65 years, (4) not having visited a neurologist for headache within 2 years prior to consultation, and (5) waiting time from referral to consultation 4 months or less.

A nurse welcomed the participants at the entrance of the neurological department at the Tromsø University Hospital, checked the patient's self-administered prefilled forms and participation consent and called the randomization administrator at the hospital (Centre for Quality Improvement and Development). Participants were block-randomized by using an Rnd function in Access (Microsoft Corp), and thereafter guided to an examination room for face-to-face consultation (traditional group) or to the video conference room located next to the department (video group). Video consultations were performed by using a video conference system including a C40 Integrator package (Cisco Systems Inc) with dual display option and Touch Control Device for C Series, C40 Integrator Multisite (Cisco Systems Inc), Precision HD 1080*P* 12× camera (Cisco Systems Inc), X551S 55" LED (Sharp NEC Display Solutions of America Inc) monitor, ceiling microphones (Audio-Technica Inc), and JBL LSR2325P (Harman International Industries) active speakers installed in the video conference room providing 2-way video and audio communication between patient and specialist. The neurologists consulted the patients from 2 other offices via a EX60 unit (Cisco Systems Inc) with an in-touch panel. Traditional face-to-face consultations were performed in the same offices. The study nurse confirmed that the visual and audio devices worked and informed the patients about the location of the web camera and microphone and where to sit. The nurse also provided a short training and assured optimal communication with the specialist. Two experienced neurologists (KIM and SIB) performed clinical consultations without neurological examinations but with additional checklists for inclusion criteria, diagnostic classification, and a standardized interview which were developed before the trial without further development during the trial. Further details are published elsewhere [19].

Data Collection and End Points

Data were obtained by structured interview at baseline and via questionnaire at 1-year follow-up. The prefilled forms included a Headache Impact Test–6 (HIT-6) measuring 6 items of headache impact (pain, social, role and cognitive functioning, vitality, and psychological distress). Every question was answered by never, rarely, sometimes, very often, or always, and each answer scored 6, 8, 10, 11, or 13 points, respectively [20]. Pain intensity using a horizontal visual analog scale (VAS)

ranging from 0 to 10 (0 = no pain, 10 = worst possible pain) was used in conjunction with HIT-6 [21]. Clinical and headache characteristics including comorbidity and diagnosis according to International Classification of Headache Disorders–2 [4] were recorded. Also, inaccurate headache diagnosis (ie, diagnostic disagreement between specialist and diagnoses reported in the electronic referral letter) were registered. The follow-up questionnaire recording demographics, clinical and headache characteristics, and end point variables was sent to the patients (with a reminder 2 weeks later to the nonresponders) either through an online survey service (Questback) or by postal letter. To classify chronic headache, the patients responded to number of headache days per month for the last 3 months.

Secondary end points were recorded from the patients' registration form as follows: "Where you satisfied with the consultation?" (yes or no), "Have you consulted your GP for headache after the specialist consultation?" (yes or no), and "Number of headache consultations with GP after the specialist consultation." Also, use of painkillers, triptans, and preventive headache drugs used in the last month were recorded.

Ethics

Oral and written consent were obtained from all participants before study entrance. The Norwegian National Committee for Medical and Health Research Ethics approved the study (number 2009/1430/REK).

Statistical Analyses

Data were analyzed with SPSS (version 27, IBM Corp). Descriptive variables are compared between the randomized groups and presented as mean and standard deviation or median and interquartile range in skewed distributed data (number of GP consultations). Consequently, comparisons between groups were analyzed by independent Student *t* test or Mann-Whitney *U* test, respectively; all 2-sided with $P<.05$ selected as level of statistical significance. Categorical variables were presented as numbers and percentages while groups were compared by using chi-square tests.

Results

Patients' characteristics were similar for both video and traditional consultations in all aspects including education and headache characteristics except for younger age in the video group (Table 1). A majority (174/233, 74.7%) had migraine as primary headache, and about one-third (79/233, 33.9%) used analgesic drugs and/or triptans or other headache-specific medication (Table 1). Consultation duration was shorter in the video group (Table 1). Inaccurate headache diagnosis (diagnostic disagreement between specialist and diagnoses reported in the electronic referral letter) are presented in Table 2. Comparisons between the video group and the face-to-face group were insignificant with respect to renewed headache diagnosis and preventive treatment given by the specialist (Table 2). The specialist prescribed preventive medication to 50% to 70% of the patients, and the group with chronic headache remission was treated similarly regardless of consultation form in that respect (Table 2). The main outcome was 43.0% (37/86) 1-year chronic headache remission rate in the video group compared to 39.5% (30/76) in the traditional group ($P=.38$; Table 3). Number and frequency of patients satisfied with consultations were 86.5% (32/37; video) and 93.3% (28/30; face-to-face; $P=.25$; Table 3). GP consultations (numbers and frequencies of patients and median numbers of consultations) are presented in Table 3. More patients treated traditionally (16/30, 53.3%) reported that they had consulted a GP for headache in the follow-up period (Table 3). No end points were otherwise statistically significantly different between the 2 groups. Neither were there any differences in changes in the HIT-6 and VAS scores from baseline to 1-year assessment between the 2 groups (Table 2).

Table 1. Clinical characteristics in randomized groups of patients referred to specialist for chronic headache consulted by video or traditionally.

Characteristic	Chronic headache at baseline			Remission from chronic headache at 12 months		
	Video (n=120)	Face-to-face (n=113)	*P* value	Video (n=37)	Face-to-face (n=30)	*P* value
One-year response, n (%)	86 (71.7)	76 (67.3)	.52	__[a]	—	—
Females, n (%)	86 (71.7)	84 (74.3)	.66	30 (72.1)	21 (70.0)	.39
Age (years), mean (SD)	35.2 (12.8)	40.0 (13.7)	.006	38.3(12.4)	41.2 (14.6)	.38
Education (years), mean (SD)	13.2 (2.9)	14.0 (3.1)	.07	13.3 (2.7)	13.9 (3.0)	.42
Sick leave (headache, weeks), n (%)	42 (35.0)	43 (38.1)	.68	11 (29.7)	12 (40.0)	.58
Waiting time to specialist (days), mean (SD)	59.0 (29.0)	55.2 (26.1)	.29	58.7 (25.5)	46.9 (23.5)	.06
Consultation duration (minutes), mean (SD)	40.2 (9.8)	46.5 (13.0)	<.001	41.0 (8.1)	45.8 (8.8)	.02
BMI (mg/m^2), mean (SD)	27.1 (5.5)	26.9 (5.7)	.79	27.8 (4.5)	28.6 (7.5)	.35
Obesity, BMI ≥30, n (%)	31 (25.8)	29 (25.7)	>.99	27 (73.0)	20 (66.7)	.58
Without comorbidity, n (%)	62 (51.7)	52 (46.0)	.54	18 (48.6)	13 (43.3)	.81
Chronic neck pain, n (%)	56 (46.7)	57 (50.4)	.60	20 (54.1)	14 (46.7)	.63
Insomnia, n (%)	80 (66.7)	72 (63.7)	.68	9 (24.3)	10 (33.3)	.43
Hypertension, n (%)	11 (9.2)	17 (15.0)	.23	5 (13.5)	4 (13.3)	>.99
Age at headache onset (years), mean (SD)	24.4 (14.3)	27.7 (14.7)	.09	26.1 (15.3)	30.2 (15.8)	.29
Headache duration (years), mean (SD)	12.2 (12.8)	13.6 (14.6)	.35	13.2 (13.2)	15.3 (16.0)	.58
Chronic headache subtype[b], n (%)						
Migraine	90 (75.0)	84 (74.3)	>.99	31 (83.8)	23 (76.6)	.79
Tension-type	23 (19.2)	28 (24.8)	.34	6 (16.2)	6 (20.0)	—
Other	7 (5.8)	1 (0.9)	—	0 (0)	1 (3.3)	—
Medication ≥15 days/month[c], n (%)	39 (32.5)	40 (35.4)	.68	8 (21.6)	3 (10.0)	—

[a]Not applicable.

[b]Most prominent headache subtype given by specialist.

[c]Use of painkillers and/or triptans ≥15 days per month last 3 month.

Table 2. Diagnostic changes and preventive chronic headache treatment given by neurologist. Comparisons between groups of patients randomized to either video or traditional consultations.

Variable	Persistent chronic headache at 12 months			Remission from chronic headache at 12 months		
	Video (n=49)	Face-to-face (n=46)	*P* value	Video (n=37)	Face-to-face (n=30)	*P* value
New headache diagnosis, n (%)	15 (30.6)	9 (19.6)	.25	9 (24.3)	9 (30.0)	.78
Preventive treatment, n (%)	26 (53.1)	29 (63.0)	.41	26 (70.3)	21 (70.0)	>.99
Antihypertensive	9 (18.4)	5 (10.9)	__[a]	9 (24.3)	3 (10.0)	—
Antiepileptic	6 (12.2)	7 (15.1)	—	7 (18.8)	6 (20.0)	—
Antidepressant	11 (22.5)	17 (37.0)	—	10 (27.0)	12 (40.0)	—
Triptans, n (%)	19 (38.8)	14 (30.4)	.55	11 (29.7)	6 (20.0)	.41

[a]Not applicable.

Table 3. Remission rates from chronic headache (primary end point), patient's satisfaction with consultation and general practitioner consultations (secondary end points), headache-related symptoms, and therapy in the 12 months after specialist consultation. Patients randomized to either video or traditional consultations.

Variables	Persistent chronic headache at 12 months			Remission from chronic headache at 12 months		
	Video (n=49)	Face-to-face (n=46)	*P* value	Video (n=37)	Face-to-face (n=30)	*P* value
Remission rate from CH[a] (%)[b]	—[c]	—	—	37/86 (43.0)	30/76 (39.5)	.38
Persistent CH (%)[b]	49/86 (57.0)	46/76 (60.5)	.38	—	—	—
Patient satisfaction with consultation, n (%)	42 (85.7)	41 (89.1)	.41	32 (86.5)	28 (93.3)	.25
GP[d] consultations, n (%)	29 (59.2)	20 (43.5)	.12	11 (29.7)	16 (53.3)	.03
GP consultations, median (IQR range)	2 (0-11)	1 (0-11)	.04	1 (0-13)	1.5 (0-15)	.19
HIT[e] -6, baseline, mean (SD)	64.0 (5.6)	64.9 (4.6)	.38	66.0 (3.9)	63.7 (6.4)	.09
HIT-6 after 1 year, mean (SD)	58.3 (8.8)	61.6 (7.8)	.10	59.9 (10.5)	59.2 (8.2)	.98
ΔHIT-6, mean (SD)	5.7 (9.3)	3.3 (8.7)	.08	5.5 (12.4)	5.8 (9.6)	.92
VAS[f], baseline, mean (SD)	6.9 (2.3)	6.9 (2.1)	.95	7.0 (2.1)	6.8 (2.0)	.66
VAS after 1 year, mean (SD)	5.2 (2.8)	6.6 (2.0)	.02	5.3 (2.8)	5.2 (3.2)	.94
ΔVAS, mean (SD)	1.7 (3.8)	0.3 (3.5)	.01	1.7 (3.3)	1.6 (3.5)	.80
Analgesic use, n (%)	38 (77.6)	39 (84.8)	.44	34 (91.9)	29 (96.7)	.62
Medication ≥15 days/month[g], n (%)	27 (55.1)	23 (50.0)	.48	8 (21.6)	3 (10.0)	—

[a]CH: chronic headache.

[b]Calculated by using response rates (per protocol analyses) as reference.

[c]Not applicable.

[d]GP: general practitioner.

[e]HIT: headache impact test.

[f]VAS: visual analog scale.

[g]Use of painkillers and/or triptans ≥15 days per month last 3 months.

When taken data from the groups together (pooled data), the comparisons between baseline and status after 12 months were as follows: remission rate from chronic headache was 41.4% (67/162) and numbers visiting GPs were 30.2% (49/162) of those with persisting chronic headache and 40.3% (27/67) in the chronic headache remission group (*P*=.41). Median numbers of GP consultations were 1.0 (IQR 0-15) and 2.0 (IQR 0-11), respectively (*P*=.25). The rate of participants using analgesic medication or triptans ≥15 days per month declined from 52.6% (50/95) to 16.4% (11/67) 1-year post consultation.

Discussion

Principal Findings

By managing new referred chronic headache patients at a secondary neurological center, the 1-year results from this post hoc RCT showed that consulting a neurological specialist by using video were equivalent to face-to-face consultations. Thus, we found no significant differences in remission rate from chronic headache, patient satisfaction with consultation, or GP visits due to headache conducted in the 1-year follow-up period. This study provides evidence to support specialist video consultations as a good alternative to face-to-face consultations in treating patients with chronic headache.

Comparison With Prior Work

There are no previous studies comparing consultation forms in treating chronic headache by a specialist, but in an earlier RCT the group of chronic headache patients randomized to an internet-delivered self-managing relaxation program (n=39) improved by 47% on measures of self-reported headache symptoms compared to an equivalent control group recruited from the waiting list with symptom monitoring only [22]. That study documents the usefulness of communicating via electronic devices as an alternative to face-to-face consultations in treating difficult headache such as chronic headache in line with this study. Likewise, the 1-year treatment response is comparable between the studies (47% vs 43%) despite different treatment methods and outcomes [22]. A smaller RCT by Friedman et al [23] randomized 18 patients with severe migraine to video consultations and 12 to in-office visits in a tertiary headache center. Improvement in headache burden and number of headache days were not different between the groups, and the authors concluded that video consultations were as effective as in-office visits. Furthermore, the consultation time was shorter in the telemedicine cohort as in this study (Table 1) indicating that telemedicine is effective for physicians in treating difficult headache [23].

In our study, approximately 40% of the chronic headache patients had remitted 1-year postconsultation while about 60% persisted with chronic headache. This rate of remission is somewhat lower than a previous longitudinal study that showed a 40% persistent rate at 1 year and 25% at 2-year follow-up [24]. Medication overuse was associated with chronicity in that study, which is also indicated here, as the rate of participants using analgesic medication or triptans ≥15 days per month declined from 52.6% to 16.4%. Similar findings are also demonstrated in population-based studies [25]. RCTs as a method to investigate different neurological outpatient management are in general few [26,27], but use of telemedicine was equivalent to face-to-face consultations with specialist as far as number of consultations [28]. Teleneurology is nevertheless widely used in clinical practice [29-31] with favorable results from the patient perspective (time- and money-saving, communication, perceiving good care, and future preference) [31,32]. From a specialist point of view (n=135 specialists), headache and follow-up consultations were well suited for telemedicine [33].

In general, RCTs in eHealth are few despite occurrence of the COVID-19 pandemic situation, which has demonstrated a need for more evidence-based knowledge about the use of digital health technology in evaluating treatment effect, safety, and other aspects of patient management [34-37]. Patient education programs [38,39]; evaluation of psychological distress using teletechnology in diabetes [40]; use of mobile in suboptimal health [41], surgery care, and follow-up [42-44]; aphasia [44]; HIV consultations [38]; cancer symptom monitoring [45]; motor and cognitive function in stroke [46]; and COVID-19 follow-up [47] are areas where RCTs are used. Moreover, this study agrees with previous telemedicine RCTs in the same area reporting positive outcomes in treating diabetic foot ulcer [48] and in a follow-up study of orthopedic patients [49]. Thus, the RCT design is the main advantage of this study, especially since it is the first one to compare consultation forms with specialist in chronic headache where one treatment arm is based on teletechnology.

Limitations

This post hoc study containing a 53% sample of the original cohort of headache sufferers may be prone to statistical type 2 failure due to risk of underpowered sample size, although the video and traditional consultation groups were similar with respect to group sizes and most of the social and clinical characteristics reflecting a design resistant to selection bias. Moreover, such a study lacks a placebo group and blinding, which would have optimized the evidence further. Awareness of the fact that this study compares different consultation forms and not specific treatment options should be emphasized. Interim analyses comparing additional clinical information between patient groups within the 1-year follow-up period might extend the knowledge about patient experiences with video consultations and should be performed in future studies. Additionally, consecutively including patients from clinical practice and a relatively low dropout rate accounts for acceptable generalizability.

Conclusions

This RCT of video consultations for new referrals of chronic headache patients demonstrated that chronic headache remission rate, patient satisfaction with specialist consultation, and GP consultations for headache performed during follow-up were equivalent between the video group and the face-to-face group. This study adds to the documentation of eHealth in consulting headache patients by specialist.

Acknowledgments

Jorun Willumsen, Anna-Kirsti Kvitnes, Marlen Lauritsen, Nora Bekkelund, Karin Flatekval Eines, Torill Erdahl, Marianne Røst, and Grethe Berg Johnsen are acknowledged for skillful help with patient logistics and data acquisition. The publication charges for this article have been funded by a grant from the publication fund of the Arctic University of Norway. The Northern Norway Regional Health Authority funded this study.

Authors' Contributions

SIB contributed to the conception and design, data collection, statistical analysis, and interpretation of data, wrote the article, and approved the final version. KIM contributed with data collection, revision of the manuscript, and approval of the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 239 KB - jmir_v23i12e30151_app1.pdf]

References

1. Natoli JL, Manack A, Dean B, Butler Q, Turkel CC, Stovner L, et al. Global prevalence of chronic migraine: a systematic review. Cephalalgia 2010 May;30(5):599-609. [doi: 10.1111/j.1468-2982.2009.01941.x] [Medline: 19614702]
2. Schwartz BS, Stewart WF, Simon D, Lipton RB. Epidemiology of tension-type headache. JAMA 1998 Feb 04;279(5):381-383. [doi: 10.1001/jama.279.5.381] [Medline: 9459472]

3. Buse DC, Manack AN, Fanning KM, Serrano D, Reed ML, Turkel CC, et al. Chronic migraine prevalence, disability, and sociodemographic factors: results from the American Migraine Prevalence and Prevention Study. Headache 2012;52(10):1456-1470. [doi: 10.1111/j.1526-4610.2012.02223.x] [Medline: 22830411]
4. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition (beta version). Cephalalgia 2013 Jul;33(9):629-808. [doi: 10.1177/0333102413485658] [Medline: 23771276]
5. Reuter U. GBD 2016: still no improvement in the burden of migraine. Lancet Neurol 2018 Nov;17(11):929-930 [FREE Full text] [doi: 10.1016/S1474-4422(18)30360-0] [Medline: 30353862]
6. GBD 2016 Headache Collaborators. Global, regional, and national burden of migraine and tension-type headache, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2018 Nov;17(11):954-976 [FREE Full text] [doi: 10.1016/S1474-4422(18)30322-3] [Medline: 30353868]
7. Karimi L, Wijeratne T, Crewther SG, Evans AE, Ebaid D, Khalil H. The migraine-anxiety comorbidity among migraineurs: a systematic review. Front Neurol 2020;11:613372 [FREE Full text] [doi: 10.3389/fneur.2020.613372] [Medline: 33536997]
8. Scher AI, Buse DC, Fanning KM, Kelly AM, Franznick DA, Adams AM, et al. Comorbid pain and migraine chronicity: The Chronic Migraine Epidemiology and Outcomes Study. Neurology 2017 Aug 01;89(5):461-468 [FREE Full text] [doi: 10.1212/WNL.0000000000004177] [Medline: 28679597]
9. Messali A, Sanderson JC, Blumenfeld AM, Goadsby PJ, Buse DC, Varon SF, et al. Direct and indirect costs of chronic and episodic migraine in the united states: a web-based survey. Headache 2016 Feb;56(2):306-322. [doi: 10.1111/head.12755] [Medline: 26833083]
10. Von Korff M, Stewart WF, Simon DJ, Lipton RB. Migraine and reduced work performance: a population-based diary study. Neurology 1998 Jun;50(6):1741-1745. [doi: 10.1212/wnl.50.6.1741] [Medline: 9633720]
11. Potter R, Probyn K, Bernstein C, Pincus T, Underwood M, Matharu M. Diagnostic and classification tools for chronic headache disorders: a systematic review. Cephalalgia 2019 May;39(6):761-784 [FREE Full text] [doi: 10.1177/0333102418806864] [Medline: 30335472]
12. Cevoli S, D'Amico D, Martelletti P, Valguarnera F, Del Bene E, De Simone R, et al. Underdiagnosis and undertreatment of migraine in Italy: a survey of patients attending for the first time 10 headache centres. Cephalalgia 2009 Dec;29(12):1285-1293. [doi: 10.1111/j.1468-2982.2009.01874.x] [Medline: 19438916]
13. Dodick DW, Loder EW, Manack Adams A, Buse DC, Fanning KM, Reed ML, et al. Assessing barriers to chronic migraine consultation, diagnosis, and treatment: results from the Chronic Migraine Epidemiology and Outcomes (CaMEO) Study. Headache 2016 May;56(5):821-834 [FREE Full text] [doi: 10.1111/head.12774] [Medline: 27143127]
14. Katsarava Z, Mania M, Lampl C, Herberhold J, Steiner TJ. Poor medical care for people with migraine in Europe: evidence from the Eurolight study. J Headache Pain 2018 Feb 01;19(1):10 [FREE Full text] [doi: 10.1186/s10194-018-0839-1] [Medline: 29392600]
15. Ryvlin P, Skorobogatykh K, Negro A, Sanchez-De La Rosa R, Israel-Willner H, Sundal C, et al. Current clinical practice in disabling and chronic migraine in the primary care setting: results from the European My-LIFE anamnesis survey. BMC Neurol 2021 Jan 04;21(1):1 [FREE Full text] [doi: 10.1186/s12883-020-02014-6] [Medline: 33390161]
16. Tassorelli C, Farm I, Kettinen H, de la Torre ER, Stretenovic S, Thomas W, et al. Access to care—an unmet need in headache management? J Headache Pain 2014;15:20 [FREE Full text] [doi: 10.1186/1129-2377-15-20] [Medline: 24742114]
17. Müller KI, Alstadhaug KB, Bekkelund SI. Acceptability, feasibility, and cost of telemedicine for nonacute headaches: a randomized study comparing video and traditional consultations. J Med Internet Res 2016 May 30;18(5):e140 [FREE Full text] [doi: 10.2196/jmir.5221] [Medline: 27241876]
18. 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]
19. Müller KI, Alstadhaug KB, Bekkelund SI. A randomized trial of telemedicine efficacy and safety for nonacute headaches. Neurology 2017 Jul 11;89(2):153-162 [FREE Full text] [doi: 10.1212/WNL.0000000000004085] [Medline: 28615434]
20. Yang M, Rendas-Baum R, Varon SF, Kosinski M. Validation of the Headache Impact Test (HIT-6™) across episodic and chronic migraine. Cephalalgia 2011 Feb;31(3):357-367 [FREE Full text] [doi: 10.1177/0333102410379890] [Medline: 20819842]
21. Lundqvist C, Benth JS, Grande RB, Aaseth K, Russell MB. A vertical VAS is a valid instrument for monitoring headache pain intensity. Cephalalgia 2009 Oct;29(10):1034-1041 [FREE Full text] [doi: 10.1111/j.1468-2982.2008.01833.x] [Medline: 19735531]
22. Devineni T, Blanchard EB. A randomized controlled trial of an internet-based treatment for chronic headache. Behav Res Ther 2005 Mar;43(3):277-292. [doi: 10.1016/j.brat.2004.01.008] [Medline: 15680926]
23. Friedman DI, Rajan B, Seidmann A. A randomized trial of telemedicine for migraine management. Cephalalgia 2019 Oct;39(12):1577-1585. [doi: 10.1177/0333102419868250] [Medline: 31450969]
24. Wang S, Fuh J, Lu S, Juang K. Outcomes and predictors of chronic daily headache in adolescents: a 2-year longitudinal study. Neurology 2007 Feb 20;68(8):591-596. [doi: 10.1212/01.wnl.0000252800.82704.62] [Medline: 17182975]

25. Hagen K, Kristoffersen ES, Winsvold BS, Stovner LJ, Zwart J. Remission of chronic headache: an 11-year follow-up study. Data from the Nord-Trøndelag Health Surveys 1995-1997 and 2006-2008. Cephalalgia 2018 Dec;38(14):2026-2034. [doi: 10.1177/0333102418769940] [Medline: 29629599]
26. Wechsler LR. Advantages and limitations of teleneurology. JAMA Neurol 2015 Mar;72(3):349-354. [doi: 10.1001/jamaneurol.2014.3844] [Medline: 25580942]
27. Wilkinson JR, Spindler M, Wood SM, Marcus SC, Weintraub D, Morley JF, et al. High patient satisfaction with telehealth in Parkinson disease: a randomized controlled study. Neurol Clin Pract 2016 Jun;6(3):241-251 [FREE Full text] [doi: 10.1212/CPJ.0000000000000252] [Medline: 27347441]
28. Chua R, Craig J, Esmonde T, Wootton R, Patterson V. Telemedicine for new neurological outpatients: putting a randomized controlled trial in the context of everyday practice. J Telemed Telecare 2002;8(5):270-273. [doi: 10.1177/1357633X0200800505] [Medline: 12396855]
29. 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]
30. Yurkiewicz IR, Lappan CM, Neely ET, Hesselbrock RR, Girard PD, Alphonso AL, et al. Outcomes from a US military neurology and traumatic brain injury telemedicine program. Neurology 2012 Sep 18;79(12):1237-1243. [doi: 10.1212/WNL.0b013e31826aac33] [Medline: 22955133]
31. Davis LE, Coleman J, Harnar J, King MK. Teleneurology: successful delivery of chronic neurologic care to 354 patients living remotely in a rural state. Telemed J E Health 2014 May;20(5):473-477. [doi: 10.1089/tmj.2013.0217] [Medline: 24617919]
32. 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]
33. Kristoffersen ES, Sandset EC, Winsvold BS, Faiz KW, Storstein AM. Experiences of telemedicine in neurological out-patient clinics during the COVID-19 pandemic. Ann Clin Transl Neurol 2021 Feb;8(2):440-447 [FREE Full text] [doi: 10.1002/acn3.51293] [Medline: 33377609]
34. Gunasekeran DV, Tseng RMWW, Tham Y, Wong TY. Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. NPJ Digit Med 2021 Feb 26;4(1):40 [FREE Full text] [doi: 10.1038/s41746-021-00412-9] [Medline: 33637833]
35. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
36. Svendsen MT, Tiedemann SN, Andersen KE. Pros and cons of eHealth: a systematic review of the literature and observations in Denmark. SAGE Open Med 2021;9:20503121211016179 [FREE Full text] [doi: 10.1177/20503121211016179] [Medline: 34046178]
37. Angus-Leppan H, Guiloff AE, Benson K, Guiloff RJ. Navigating migraine care through the COVID-19 pandemic: an update. J Neurol 2021 May 17:1-8 [FREE Full text] [doi: 10.1007/s00415-021-10610-w] [Medline: 34002281]
38. Hickey MD, Sergi F, Zhang K, Spinelli MA, Black D, Sola C, et al. Pragmatic randomized trial of a pre-visit intervention to improve the quality of telemedicine visits for vulnerable patients living with HIV. J Telemed Telecare 2020 Dec 20:1357633X20976036. [doi: 10.1177/1357633X20976036] [Medline: 33342328]
39. Allida S, Du H, Xu X, Prichard R, Chang S, Hickman LD, et al. mHealth education interventions in heart failure. Cochrane Database Syst Rev 2020 Jul 02;7:CD011845 [FREE Full text] [doi: 10.1002/14651858.CD011845.pub2] [Medline: 32613635]
40. Alessi J, de Oliveira GB, Franco DW, Becker AS, Knijnik CP, Kobe GL, et al. Telehealth strategy to mitigate the negative psychological impact of the COVID-19 pandemic on type 2 diabetes: a randomized controlled trial. Acta Diabetol 2021 Jul;58(7):899-909. [doi: 10.1007/s00592-021-01690-1] [Medline: 33723649]
41. Baek Y, Jeong K, Lee S, Kim H, Seo B, Jin H. Feasibility and effectiveness of assessing subhealth using a mobile health management app (MibyeongBogam) in early middle-aged koreans: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 19;9(8):e27455 [FREE Full text] [doi: 10.2196/27455] [Medline: 34420922]
42. Binder EF, Christensen JC, Stevens-Lapsley J, Bartley J, Berry SD, Dobs AS, et al. A multi-center trial of exercise and testosterone therapy in women after hip fracture: design, methods and impact of the COVID-19 pandemic. Contemp Clin Trials 2021 May;104:106356. [doi: 10.1016/j.cct.2021.106356] [Medline: 33716173]
43. Kolcun JPG, Ryu WHA, Traynelis VC. Systematic review of telemedicine in spine surgery. J Neurosurg Spine 2020 Oct 30:1-10. [doi: 10.3171/2020.6.SPINE20863] [Medline: 33126219]
44. Harkey K, Kaiser N, Zhao J, Hetherington T, Gutnik B, Matthews BD, et al. Postdischarge virtual visits for low-risk surgeries: a randomized noninferiority clinical trial. JAMA Surg 2021 Mar 01;156(3):221-228. [doi: 10.1001/jamasurg.2020.6265] [Medline: 33439221]
45. Maguire R, McCann L, Kotronoulas G, Kearney N, Ream E, Armes J, et al. Real time remote symptom monitoring during chemotherapy for cancer: European multicentre randomised controlled trial (eSMART). BMJ 2021 Jul 21;374:n1647 [FREE Full text] [doi: 10.1136/bmj.n1647] [Medline: 34289996]

46. Ozen S, Senlikci HB, Guzel S, Yemisci OU. Computer game assisted task specific exercises in the treatment of motor and cognitive function and quality of life in stroke: a randomized control study. J Stroke Cerebrovasc Dis 2021 Sep;30(9):105991. [doi: 10.1016/j.jstrokecerebrovasdis.2021.105991] [Medline: 34293643]
47. Rodriguez-Blanco C, Gonzalez-Gerez JJ, Bernal-Utrera C, Anarte-Lazo E, Perez-Ale M, Saavedra-Hernandez M. Short-term effects of a conditioning telerehabilitation program in confined patients affected by COVID-19 in the acute phase. a pilot randomized controlled trial. Medicina (Kaunas) 2021 Jul 03;57(7):1 [FREE Full text] [doi: 10.3390/medicina57070684] [Medline: 34356965]
48. Iversen MM, Igland J, Smith-Strøm H, Østbye T, Tell GS, Skeie S, et al. Effect of a telemedicine intervention for diabetes-related foot ulcers on health, well-being and quality of life: secondary outcomes from a cluster randomized controlled trial (DiaFOTo). BMC Endocr Disord 2020 Oct 21;20(1):157 [FREE Full text] [doi: 10.1186/s12902-020-00637-x] [Medline: 33087074]
49. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient satisfaction with remote orthopaedic consultation by using telemedicine: a randomised controlled trial. J Telemed Telecare 2018 Jan 01:1357633X18783921. [doi: 10.1177/1357633X18783921] [Medline: 29973130]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online TeleHealth
GP: general practitioner
HIT-6: Headache Impact Test–6
RCT: randomized controlled trial
VAS: visual analog scale

Edited by R Kukafka; submitted 04.05.21; peer-reviewed by K Johnson, D Gunasekeran, A Hidki; comments to author 21.07.21; revised version received 04.09.21; accepted 12.09.21; published 13.12.21.

Please cite as:
Bekkelund SI, Müller KI
One-Year Remission Rate of Chronic Headache Comparing Video and Face-to-Face Consultations by Neurologist: Randomized Controlled Trial
J Med Internet Res 2021;23(12):e30151
URL: https://www.jmir.org/2021/12/e30151
doi:10.2196/30151
PMID:34898455

Original Paper

Use of the McIsaac Score to Predict Group A Streptococcal Pharyngitis in Outpatient Nurse Phone Triage and Electronic Visits Compared With In-Person Visits: Retrospective Observational Study

Jennifer L Pecina[1], MD; Leah M Nigon[2], RN, MSN; Kristine S Penza[3], APRN, CNP; Martha A Murray[3], APRN, CNP; Beckie J Kronebusch[2], RN, MSN; Nathaniel E Miller[1], MD; Teresa B Jensen[1], MD

[1]Department of Family Medicine, Mayo Clinic, Rochester, MN, United States

[2]Department of Nursing, Mayo Clinic, Rochester, MN, United States

[3]Mayo Clinic Express Care, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Jennifer L Pecina, MD
Department of Family Medicine
Mayo Clinic
200 1st street SW
Rochester, MN
United States
Phone: 1 507 284 2511
Email: pecina.jennifer@mayo.edu

Abstract

Background: The McIsaac criteria are a validated scoring system used to determine the likelihood of an acute sore throat being caused by group A streptococcus (GAS) to stratify patients who need strep testing.

Objective: We aim to compare McIsaac criteria obtained during face-to-face (f2f) and non-f2f encounters.

Methods: This retrospective study compared the percentage of positive GAS tests by McIsaac score for scores calculated during nurse protocol phone encounters, e-visits (electronic visits), and in person f2f clinic visits.

Results: There was no difference in percentages of positive strep tests between encounter types for any of the McIsaac scores. There were significantly more phone and e-visit encounters with any missing score components compared with f2f visits. For individual score components, there were significantly fewer e-visits missing fever and cough information compared with phone encounters and f2f encounters. F2f encounters were significantly less likely to be missing descriptions of tonsils and lymphadenopathy compared with phone and e-visit encounters. McIsaac scores of 4 had positive GAS rates of 55% to 68% across encounter types. There were 4 encounters not missing any score components with a McIsaac score of 0. None of these 4 encounters had a positive GAS test.

Conclusions: McIsaac scores of 4 collected during non-f2f care could be used to consider empiric treatment for GAS without testing if significant barriers to testing exist such as the COVID-19 pandemic or geographic barriers. Future studies should evaluate further whether non-f2f encounters with McIsaac scores of 0 can be safely excluded from GAS testing.

(J Med Internet Res 2021;23(12):e25899) doi:10.2196/25899

KEYWORDS

strep pharygitis; e-visit; electronic visit; telemedicine; telecare; virtual visit; McIssac score; nurse phone triage; scoring system; sore throat; group A streptococcus; telehealth; nurse; phone; triage

Introduction

The McIsaac (modified Centor) scoring system is a validated tool used to determine the likelihood of an acute sore throat being from group A streptococcus (GAS) [1-3]. Criterion include patient age, absence of cough, fever, tonsillar swelling or exudates, and the presence of anterior cervical lymphadenopathy, with the last 2 criteria requiring a physical

examination [1]. Even with 4 of the criteria present, the likelihood of an infection with strep is around 60% [4], so best practice dictates patients should be swabbed for GAS prior to initiating treatment with an antibiotic. Current guidelines also suggest against GAS testing in patients with a McIsaac score of less than 3 [5,6]. The most recent Infectious Diseases Society of America guidelines recommend considering testing unless other features strongly suggest a viral etiology [7].

The McIsaac criteria were based on clinician assessment and were not intended to be based on patient/caregiver report. The ability of patients and parents to assess tonsils and lymph nodes is not well studied. Minimal evidence exists in the literature related to use of the McIsaac criteria being calculated based on patient/caregiver report in a non–face-to-face (non-f2f) encounter. A previous study showed that adult patients may over- or underreport physical exam findings compared with clinicians [8]. Another study comparing patients or caregivers to physicians showed moderate to substantial agreement for 3 of 4 key pharyngitis signs and symptoms [9].

In 2017, our institution began using the McIsaac scoring criteria for non-f2f nurse phone triage and e-visit encounters. This scoring system was chosen due to it being validated and having relatively few components needing a response from patients/caregivers. Both phone triage and e-visit encounters relied on patients/caregivers to report on historical symptoms of cough and fever as well as physical exam findings of enlarged tonsils, tonsillar exudate, and anterior cervical lymphadenopathy. Both phone triage encounters and e-visits also allowed patients/caregivers to respond with “I don’t know” for any of the McIsaac criteria with the exception of e-visits not allowing an “I don’t know” response for cough. e-Visits were text-based asynchronous visits that did not require photo or video capability. To complete an e-visit, patients/caregivers completed an online questionnaire via a secure patient portal which was then reviewed by family medicine nurse practitioners and physician assistants. Both e-visit and phone triage end points recommended reporting for strep testing if the McIsaac score was 3 or higher. However, some patients were also referred for a f2f visit or strep testing with scores less than 3 either by clinician judgement or, at times, due to patient/caregiver desire. Additionally, triage nurses sent patients for strep testing if the patient or caregiver was not able to assess any components of the score.

The primary aim of this study was to evaluate and compare use of the McIsaac criteria to predict positive GAS tests in e-visits, phone triage, and f2f visits. We hypothesized that there would not be a significant difference in positive GAS results by McIsaac score between the different encounter types.

Methods

This retrospective, observational study evaluated patients, ages 3 to 75 years, who had a nurse triage phone encounter, submitted an e-visit, or had a f2f visit for acute sore throat between February 23, 2017, and May 4, 2018, and had subsequent GAS testing done on the same day as the encounter. GAS testing was done using a polymerase chain reaction test. During the study time period, there were 211 e-visits with same day strep tests. All 211 e-visit encounters and a randomly selected sample of 211 records for both phone and f2f encounters were manually reviewed to determine the McIsaac score by patient/caregiver self-assessment during non-f2f encounters (nurse phone triage and e-visits) and by clinicians at f2f encounters. For f2f encounters, a score of 1 for fever was given if the provider included in their history that the patient reported a fever or the vital signs for the visit included a temperature greater than 38 °C. F2f visits without a fever at the time of the visit and with no mention in the note of the patient/caregiver reporting presence or absence of fever were counted as missing data for fever. Scores for each encounter type were compared with GAS results. We also calculated receiver operating characteristic (ROC) curves and area under the curve (AUC) for the McIsaac score for each encounter type. Patients were excluded if they had previous treatment for GAS within 30 days prior to the encounter, were currently on antibiotics for acute infection, or were younger than age 3 years or older than age 75 years. Statistical methods used are listed in the result tables. JMP PRO (version 14.1.0, SAS Institute Inc) was used to perform statistical analysis. This study was approved by the Mayo Clinic institutional review board.

Results

There was no difference in average patient age or percentage of positive strep tests between encounter types (Table 1).

There was a significantly higher percentage of patients aged 15 to 44 years and a lower percentage of patients aged 45 years and older with e-visits compared to phone call and f2f encounters. There were significantly more phone and e-visit encounters with missing score components compared with f2f visits. For individual score components, there were significantly less e-visits missing fever and cough information compared with phone call and f2f encounters. F2f encounters were significantly less likely to be missing descriptions of tonsils and lymphadenopathy compared with phone and e-visit encounters.

Percentages of positive strep tests for each McIsaac score by encounter type are shown in Table 2. There were no significant differences in percentages of positive strep tests between encounter types for any of the McIsaac scores.

We also reviewed encounters that had all components to document a McIsaac score (ie, no missing criteria). There were 342 encounters with no missing score elements (101 phone encounters, 83 e-visits and 158 f2f visits). Of these encounters, 52.1% (178/342) had a positive strep test. There were no significant differences between encounter types for percentage positive strep tests for all McIsaac scores (Table 3).

ROC AUC for e-visits was 0.62, for f2f was 0.69, and for phone encounters was 0.62.

Table 1. Patient age, positive strep tests, and missing score elements for all encounter types.

	Total (n=633)	Phone call (n=211)	e-Visit (n=211)	Face-to-face visit (n=211)	*P* value
Positive strep tests, n (%)	312 (49)	96 (45)	113 (53)	103 (49)	.25[a]
Age (years), mean	22.2	22	22.6	21.9	.86[b]
3-14, n (%)	288 (45)	97 (46)	92 (44)	99 (47)	.78[a]
15-44, n (%)	294 (46)	91 (43)	113 (53)	90 (43)	.04[a]
≥45, n (%)	51 (8)	23 (11)	6 (3)	22 (10)	.001[a]
Missing McIsaac criteria, n (%)	291 (46)	110 (52)	128 (61)	53 (25)	<.001[a]
Fever description	30 (5)	5 (17)	4 (2)	21 (10)	<.001[a]
Cough description	80 (13)	45 (21)	0	35 (17)	<.001[a]
Tonsil description	171 (27)	72 (34)	98 (46)	1 (1)	<.001[a]
Lymph node description	116 (18)	56 (26)	58 (28)	2 (1)	<.001[a]

[a]Chi-square test.

[b]Analysis of variance.

Table 2. Percentage positive strep tests by encounter type and McIsaac score for all encounters.

McIsaac score	All, % (95% CI) [n pos/n total]	Phone call, % (95% CI) [n pos/n total]	e-Visit, % (95% CI) [n pos/n total]	Face to face, % (95% CI) [n pos/n total]	*P* value[a]
0	17 (7-37) [4/23]	21 (8-48) [3/14]	0 [0/2]	14 (3-51) [1/7)	.62
1	30 (20-42) [20/66]	35 (19-55) [8/23]	33 (14-61) [4/12]	26 (14-43) [8/31]	.75
2	35 (27-44) [44/125]	31 (18-47) [11/36]	42 (28-58) [17/40}	33 (21-47) [16/49]	.50
3	54 (47-60) [119/222]	53 (42-64) [40/76]	52 (42-62) [46/88]	57 (44-69) [33/58]	.85
4	63 (56-70) [125/197]	55 (42-67) [34/62]	67 (55-77) [46/69]	68 (56-78) [45/66]	.24
s2	31.8 (25.9-38.2) [68/214]	31.8 (25.9-38/3) [68/214]	38.9 (27-52.2) [21/54]	28.7 (20.3-39) [25/87]	.43

[a]Chi-square test.

Table 3. Percentage positive strep tests by encounter type and McIsaac score for 342 encounters with no missing McIsaac criteria.

McIsaac score	All, % (95% CI) [n pos/n total]	Phone call, % (95% CI) [n pos/n total]	e-Visit, % (95% CI) [n pos/n total]	Face to face, % (95% CI) [n pos/n total]	*P* value
0	0 [0/4]	0 [0/1]	—[a]	0 [0/3]	>.99[b]
1	27 (13.7-46.1) [7/26]	50 (15-85) [2/4]	—	22.7 10-43.4) [5/22]	.28[c]
2	28.3 (17.3-42.5) [13/46]	28.6 (8.2-64.1) [2/7]	22 (6.3-54.7) [2/9}	30 (16.6-47.9) [9/30}	.90[c]
3	49.1 (40-58.4) [53/108]	48.8 (34.2-63.5) [20/41]	40 (23.4-59.3) [10/25]	54.8 (40-68.8) [23/42]	.61[c]
4	66.5 (58.8-73.3) [105/158]	58.3 (44.3-71.1) [28/48]	71 (57.6-82.1) [35/49]	68.9 (56.4-79.1) [42/61]	.35[c]
≤2	26.3 (17.7-37.1) [20/76]	33.3 (13.8-60.9) [4/12]	22.2 (6.3-54.8) [2/9]	25.4 (15.8-38.3) [14/55]	.84[c]

[a]Not applicable.

[b]Fisher exact test.

[c]Chi-square test.

Discussion

Principal Findings

The use of nursing protocols and e-visits in today's outpatient settings has improved access for patients and proven to be a patient satisfier [10-14]. Patients today want health care that fits their schedules with tests and treatments accomplished with minimal waiting [15,16]. The ability to access a triage nurse 24/7 or complete an e-visit with a response within a few hours has improved the care of outpatients and allowed more access for patients who need to be seen by a provider [10,13-15,17].

In addition, the cost of such care is greatly reduced from the usual clinic visit [11,13,14,18,19].

The validation of non-f2f visits for sore throats is important when determining the safety of this type of care. Potential ramifications of not treating GAS infections include the possibility of rheumatic fever leading to rheumatic heart disease, invasive GAS diseases, and acute glomerulonephritis [5,20-22]. GAS tonsillitis can also lead to significant discomfort [23]. Knowledge that a sore throat is due to a strep infection versus a viral cause could help reduce complications and get patients back to work or school after antibiotic treatment or home care for an upper respiratory viral infection.

Our study showed no significant difference in the percentage of positive strep tests for each McIsaac score when compared across f2f and non-f2f encounter types. Additionally, ROC AUC values were very similar for the McIsaac score across encounter types. These findings support the use of McIsaac criteria in non-f2f encounters as being comparable to use in f2f encounters. However, our study showed a high rate of positive strep tests for low McIsaac scores in all encounter types (including f2f visits) compared with the literature. Specifically, our rates of 0% to 21% positive strep tests for McIsaac scores of 0 are higher than the published values of 1% to 2.5% [1,4,24]. Similarly, for McIsaac scores of 1 our findings of a positive strep test of 26% to 35% across encounter types appears higher than the published literature of 5% to 10% [1,4,24]. This could be a result of our methodology of including patients who did not have complete data to calculate a McIsaac score and choosing to count missing data (either due to not being recorded in encounters or being answered as “I don’t know” in non-f2f encounters) as contributing 0 points to the total McIsaac score. We attempted to correct for this limitation by further analyzing only encounters with no missing data to compute a score. There were only 4 encounter types with all score components with McIsaac scores of 0, and none of these encounters had a positive strep test. There were only 26 encounters with no missing data and McIsaac scores of 1 with a positive strep rate of 22% to 50%. Another consideration for the higher percentage of positive strep tests is the presence of a carrier state since as many as 12% to 32% of children may be pharyngeal carriers of GAS [25,26]. A throat swab would be unable to differentiate whether the pharyngitis is due to acute GAS infection or viral illness in a GAS carrier [25-27].

In contrast to the above findings, the percentage of positive strep tests for McIsaac scores of 4 appears comparable to rates listed in the published literature for all encounter types (both for those not missing any score components and those with missing score components). Some previous guidelines have recommended consideration of empiric treatment for strep throat for patients with McIsaac scores of 4 or higher [24]. Our study lends support to empiric treatment of patients with McIsaac scores of 4 in non-f2f encounters, as the percentage of positive strep tests is high in this group without any significant differences between encounter types and is consistent with the published literature. This finding may be helpful in scenarios where f2f visits are challenging due to geographical barriers, reduced access to appointments in the clinics during influenza season, or even care barriers such as our current COVID-19 pandemic. Provider visits were more complete in documenting the McIsaac criteria than the phone triage or e-visits. In particular, presence or absence of tonsillar enlargement/exudate and cervical lymphadenopathy were more consistently noted in f2f visits then in non-f2f visits. However, e-visits were significantly less likely to be missing information on fever. No e-visits were missing data on the presence or absence of cough; however, this was due to the e-visit process not allowing an answer of “I don’t know” for cough. The e-visit consists of a series of questions to which the patient can answer yes, no, or “I don’t know” (for most answers), making the criteria easier to document. F2f clinic visits were not as likely as e-visits to document fever and cough. The clinic visits do not have a template for the McIsaac criteria, which most likely explains this difference in documentation.

There were no differences between encounter types in the average age of patients evaluated. Interestingly, however, patients aged 15 to 44 years more often used e-visits as their method of seeking care for a sore throat. This age group consistently uses the digital platform for health care at our institution more often than any other group. This reliance on electronic portals for health care has been the newest method to improve clinic access, reduce waiting time for patients seeking answers to health questions, and improve time spent on renewing prescriptions and allowing patients to access their health records anytime of the day or night. As this younger group ages, more electronic access to health care will need to be available.

Limitations

Prima facie, our study showed a surprisingly high percentage of positive strep tests in the lower McIsaac scores as well as an overall high rate of positive GAS tests (49%). The high percentage of positive GAS tests in our study (both for low McIsaac scores and overall) is very likely due to the retrospective methodology used. In prospective studies of the McIsaac criteria, all patients have GAS testing done regardless of their McIsaac score or other risk factors for a positive GAS test [4,28]. Due to the retrospective nature of this study, only patients who had GAS testing recommended and done as part of their clinical care were included. This methodology means there are likely many patients with low McIsaac scores who were not represented in our dataset as they would have been had this been a prospective trial. Not including these patients with low McIsaac scores in our study (as they did not have GAS testing done) likely elevated the positive rate for the low McIsaac scores and thus the overall positive rate of tests as, in a prospective trial, these patients would have had GAS tests that would be more likely to have been negative. Specifically, we anticipate that patients with low McIsaac scores who were included in our study are likely to have had a higher pretest probability of having a positive GAS test compared with patients with low scores who were not recommended for testing. This higher pretest probability is likely since these patients were recommended to have GAS testing done despite a low McIsaac score. Reasons other than a high McIsaac score that could increase the pretest probability of having a positive test (and thus leading to these patients being tested) could include a history of positive GAS tests (either due to recurrent strep

infections or carrier status) or a known strep contact. These factors would not be reflected in the McIsaac score but would lead to a higher pretest probability of testing positive for GAS and may have led to recommendations for these low McIsaac score patients to be tested, thus increasing the percentage of low McIsaac score patients with a positive GAS test. This would, in turn, elevate the overall positive GAS rate in our study. This supposition is further supported by the fact that our study has a lower percentage of patients with McIsaac scores of 0 and 1 (14% of our study population) when compared with a prospective multisite nationwide McIsaac validating study by Fine et al [4], which had 41% of their study population in the McIsaac 0 and 1 categories. A better comparison to validated studies would be comparing higher McIsaac scores of 3 or 4 where strep testing is routinely recommended. Our study found positive GAS rates of 54% and 63%, respectively, for these scores. This is closer to, although still somewhat higher than, the prospective study by Fine et al [4], where rates were 38% and 57%, respectively. Although somewhat higher, our study also included a significantly younger population (average 22.2 years) versus 34 years of age for Fine et al [4]. Given that younger patients have a higher pretest probability of a positive GAS test, this partially helps to explain the remaining discrepancy in the positivity rate [29].

Conclusions

Measures to combat the above limitations could be a prospective trial performing GAS testing in all patients (including all patients with low McIsaac scores). Methods to improve non-f2f assessment of physical exam findings could include options for patients to send in photos of their tonsils for the e-visit and phone encounters for a provider to review and provide that component of the score. Additionally, photos with examples of various degrees of tonsillar hypertrophy and exudate could be included for patients to choose which appeared similar to their own exam findings. Descriptions (and diagrams with location) of cervical lymph node enlargement could also be included in e-visits to improve patient and caregiver assessment of this score component. Both changes have in fact been instituted in new processes at our institution.

There is no difference in the percentage of positive GAS tests for each McIsaac score when comparing f2f and non-f2f care. Our study is supportive that patients with a McIsaac score of 0 and no missing score components might be candidates to safely exclude from strep testing as none of these patients in our study had a positive strep test; however, as our numbers for this category are small this should be further evaluated in future studies. Additionally, McIsaac scores of 4 or higher calculated during non-f2f care could be considered for empiric treatment of GAS pharyngitis without confirmatory testing, especially when there are significant barriers to obtaining testing.

Conflicts of Interest

None declared.

References

1. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. CMAJ 1998 Jan 13;158(1):75-83 [FREE Full text] [Medline: 9475915]
2. McIsaac WJ, Kellner J, Aufricht P, Vanjaka A, Low DE. Empirical validation of guidelines for the management of pharyngitis in children and adults. JAMA 2004 Apr 07;291(13):1587-1595. [doi: 10.1001/jama.291.13.1587] [Medline: 15069046]
3. Willis BH, Coomar D, Baragilly M. Comparison of Centor and McIsaac scores in primary care: a meta-analysis over multiple thresholds. Br J Gen Pract 2020 Mar 09;70(693):e245-e254. [doi: 10.3399/bjgp20x708833]
4. Fine AM, Nizet V, Mandl KD. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. Arch Intern Med 2012 Jun 11;172(11):847-852 [FREE Full text] [doi: 10.1001/archinternmed.2012.950] [Medline: 22566485]
5. Shulman ST, Kliegman RM. Shulman, C. In: Nelson Textbook of Pediatrics. 21st edition. Philadelphia: Elsevier; 2020.
6. Short SB, Marshall P, Miller N, Olmschenk D, Prigge K. Diagnosis and treatment of respiratory illness in children and adults. 2017. URL: https://www.icsi.org/wp-content/uploads/2019/01/RespIllness.pdf [accessed 2020-10-01]
7. Shulman S, Bisno A, Clegg H, Gerber M, Kaplan E, Lee G, Infectious Diseases Society of America. Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. Clin Infect Dis 2012 Nov 15;55(10):e86-e102 [FREE Full text] [doi: 10.1093/cid/cis629] [Medline: 22965026]
8. Xu J, Schwartz K, Monsur J, Northrup J, Neale A. Patient-clinician agreement on signs and symptoms of 'strep throat': a MetroNet study. Fam Pract 2004 Dec 01;21(6):599-604. [doi: 10.1093/fampra/cmh604] [Medline: 15528291]
9. Lindgren C, Neuman MI, Monuteaux MC, Mandl KD, Fine AM. Patient and parent-reported signs and symptoms for group A streptococcal pharyngitis. Pediatrics 2016 Jul 08;138(1):e20160317. [doi: 10.1542/peds.2016-0317] [Medline: 27279649]
10. Stamenova V, Agarwal P, Kelley L, Fujioka J, Nguyen M, Phung M, et al. Uptake and patient and provider communication modality preferences of virtual visits in primary care: a retrospective cohort study in Canada. BMJ Open 2020 Jul 06;10(7):e037064 [FREE Full text] [doi: 10.1136/bmjopen-2020-037064] [Medline: 32636284]
11. 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]

12. Li L, Georgiou A, Xiong J, Byrne M, Robinson M, Westbrook JI. Healthdirect's After Hours GP Helpline: a survey of patient satisfaction with the service and compliance with advice. Stud Health Technol Inform 2016;227:87-92. [Medline: 27440294]
13. O'Connell J, Stanley JL, Malakar CL. Satisfaction and patient outcomes of a telephone-based nurse triage service. Manag Care 2001 Jul;10(7):55-59. [Medline: 11494822]
14. Taylor L, Portnoy JM. Telemedicine for General Pediatrics. Pediatr Ann 2019 Dec;48(12):1. [doi: 10.3928/19382359-20191119-01]
15. Robb JF, Hyland MH, Goodman AD. Comparison of telemedicine versus in-person visits for persons with multiple sclerosis: a randomized crossover study of feasibility, cost, and satisfaction. Mult Scler Relat Disord 2019 Nov;36:101258. [doi: 10.1016/j.msard.2019.05.001] [Medline: 31472419]
16. Savage C, Bjessmo S, Borisenko O, Larsson H, Karlsson J, Mazzocato P. Translating 'See-and-Treat' to primary care: opening the gates does not cause a flood. Int J Qual Health Care 2019 Aug 01;31(7):30-36 [FREE Full text] [doi: 10.1093/intqhc/mzy244] [Medline: 30624735]
17. Verzantvoort NCM, Teunis T, Verheij TJM, van der Velden AW. Self-triage for acute primary care via a smartphone application: practical, safe and efficient? PLoS One 2018 Jun 26;13(6):e0199284 [FREE Full text] [doi: 10.1371/journal.pone.0199284] [Medline: 29944708]
18. Navratil-Strawn JL, Ozminkowski RJ, Hartley SK. An economic analysis of a nurse-led telephone triage service. J Telemed Telecare 2014 Jul 24;20(6):330-338. [doi: 10.1177/1357633x14545430]
19. Navratil-Strawn J, Hawkins K, Wells TS, Ozminkowski RJ, Hawkins-Koch J, Chan H, et al. Listening to the nurse pays off: an integrated Nurse HealthLine programme was associated with significant cost savings. J Nurs Manag 2014 Oct 23;22(7):837-847 [FREE Full text] [doi: 10.1111/jonm.12048] [Medline: 23607510]
20. Watkins D, Kumar R, Parks T. Epidemiology, risk factors, burden and cost of acute rheumatic fever and rheumatic heart disease. Acute Rheum Fever Rheum Heart Dis 2021:1-18. [doi: 10.1016/b978-0-323-63982-8.00001-5]
21. VanDeVoorde RG. Acute poststreptococcal glomerulonephritis: the most common acute glomerulonephritis. Pediatr Rev 2015 Jan 01;36(1):3-12. [doi: 10.1542/pir.36-1-3] [Medline: 25554106]
22. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis 2005 Nov;5(11):685-694. [doi: 10.1016/s1473-3099(05)70267-x]
23. Stjernquist-Desatnik A, Orrling A. Pharyngotonsillitis. Periodontol 2009 Feb;49:140-150 [FREE Full text] [doi: 10.1111/j.1600-0757.2008.00282.x] [Medline: 19152531]
24. Choby BA. Diagnosis and treatment of streptococcal pharyngitis. Am Fam Physician 2009 Mar 01;79(5):383-390 [FREE Full text] [Medline: 19275067]
25. Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. Pediatrics 2010 Sep 09;126(3):e557-e564. [doi: 10.1542/peds.2009-2648] [Medline: 20696723]
26. Martin JM, Green M, Barbadora KA, Wald ER. Group A streptococci among school-aged children: clinical characteristics and the carrier state. Pediatrics 2004 Nov 01;114(5):1212-1219. [doi: 10.1542/peds.2004-0133] [Medline: 15520098]
27. Gunnarsson RK, Holm SE, Söderström M. The prevalence of beta-haemolytic streptococci in throat specimens from healthy children and adults. Implications for the clinical value of throat cultures. Scand J Prim Health Care 1997 Sep 12;15(3):149-155. [doi: 10.3109/02813439709018506] [Medline: 9323783]
28. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. CMAJ 1998 Jan 13;158(1):75-83 [FREE Full text] [Medline: 9475915]
29. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. CMAJ 2000 Oct 03;163(7):811-815 [FREE Full text] [Medline: 11033707]

Abbreviations

AUC: area under the curve
f2f: face-to-face
GAS: group A streptococcus
ROC: receiver operating characteristic

Edited by R Kukafka; submitted 23.11.20; peer-reviewed by A Fine, J op den Buijs; comments to author 18.02.21; revised version received 04.03.21; accepted 24.10.21; published 20.12.21.

Please cite as:
Pecina JL, Nigon LM, Penza KS, Murray MA, Kronebusch BJ, Miller NE, Jensen TB
Use of the McIsaac Score to Predict Group A Streptococcal Pharyngitis in Outpatient Nurse Phone Triage and Electronic Visits Compared With In-Person Visits: Retrospective Observational Study
J Med Internet Res 2021;23(12):e25899
URL: https://www.jmir.org/2021/12/e25899
doi:10.2196/25899
PMID:34932016

Original Paper

The Use of Smart Speakers in Care Home Residents: Implementation Study

Katie J Edwards[1], BSc, MSc; Ray B Jones[1], PhD; Deborah Shenton[1], PhD; Toni Page[1], PhD; Inocencio Maramba[1], MD; Alison Warren[2], PhD; Fiona Fraser[2], MSc; Tanja Križaj[2], PhD; Tristan Coombe[3], RNC; Hazel Cowls[3], MSc; Arunangsu Chatterjee[1], PhD

[1]Centre for Health Technology, University of Plymouth, Plymouth, United Kingdom
[2]School of Health Professions, University of Plymouth, Plymouth, United Kingdom
[3]School of Nursing and Midwifery, University of Plymouth, Truro, United Kingdom

Corresponding Author:
Katie J Edwards, BSc, MSc
Centre for Health Technology
University of Plymouth
Desk 6, Formation Zone, Health and Wellbeing Innovation Centre
The Knowledge Spa
Plymouth, TR1 3HD
United Kingdom
Phone: 44 07432155243
Email: katie.edwards@plymouth.ac.uk

Abstract

Background: The use of smart speakers to improve well-being had been trialed in social care by others; however, we were not aware of their implementation in most care homes across a region in the Southwest of the United Kingdom. For the widespread adoption of new technology, it must be locally demonstrable and become normalized.

Objective: The aim of this study was to install smart speakers in care homes in a rural and coastal region and to explore if and how the devices were being used, the barriers to their implementation, and their potential benefits.

Methods: Email, workshops, drop-in sessions, phone, and cold calling was used to contact all 230 care homes, offering a free smart speaker and some advisory support. Care homes accepting the devices were asked to complete a feedback diary. Nonresponse rate for diary completion was high and was thus supplemented with a telephone survey.

Results: Over the course of 7 months, we installed 156 devices in 92 care homes for older people, 50 devices for people with physical or mental health needs, and 8 for others. The devices were used mainly for music but also for poetry, recipes, light controls, jokes, and video calls. Care home managers reported the benefits for the residents, including enhanced engagement with home activities, enjoyment, calming effects, and the acquisition of new skills. Implementation problems included internet connectivity, staff capacity, and skills.

Conclusions: Affordable consumer devices such as smart speakers should be installed in all care homes to benefit residents. Voice-activated technologies are easy to use and promote interaction. This study indicates that implementation in care homes was possible and that smart speakers had multifaceted benefits for residents and staff. Most care homes in this region now use smart speakers for their residents, thereby normalizing this practice.

(J Med Internet Res 2021;23(12):e26767) doi:10.2196/26767

KEYWORDS

voice-activated technology; smart speaker; care home; technology-enabled care; older people; learning disability; digital technology; consumer device; smart device

Introduction

The United Kingdom currently has 425,000 people requiring some form of residential care in over 11,000 care homes [1]. In Cornwall, two-thirds of care homes are for older people and one-third for people requiring physical and mental health care. Care homes are under pressure from an ageing population, an underresourced workforce [2], and currently from the impact of COVID-19.

Both the Social Care Institute for Excellence and the Care Quality Commission recognize the importance of technology in providing high quality care services [3,4]. For example, technology can help address loneliness through internet training and video call interventions [5].

Many projects have focused on cutting edge technologies still under development [6] rather than commercially available products that may improve quality of life now in residential care. Exceptions include the use of affordable companion robots [7]. Implementing technology use such as Skype (Microsoft Inc) in care homes has not always been straightforward [8]. Studies of smart home technology have identified desired features such as emergency help, health monitoring, and environmental control [9,10]; however, their implementation may be prevented due to cost [11].

The literature has numerous examples of "pilots" and small-scale studies and national programmes, such as the United Kingdom "delivering assisted living lifestyles at scale" project [12,13], which were set up to try to address implementation at scale conceptual frameworks on technology implementation such as Greenhalgh's Non-adoption, Abandonment, Scale-up, Spread, Sustatinability framework [14], May and Mair's Normalization Process Theory (NPT) [15], the longer standing Technology Acceptance Model [16] and the Unified Theory of Acceptance and Use of Technolgy [17], and the original work of Roger [18] all include the idea of demonstrability at scale and reaching critical mass [19]. Although the Unified Theory of Acceptance and Use of Technology 2 framework has previously been used to examine the adoption of voice-activated digital assistants with community-dwelling older adults [20], we find the NPT framework to be more appropriate for assessing and enhancing the implementation of complex interventions into routine practice, the process known as normalization. The 4 NPT components for the successful integration of interventions are coherence, cognitive participation, collective action, and reflexive monitoring.

Smart speakers became commercially available in 2014 [21] and have seen a rapid uptake with 20% of households now owning one [22]. In 2018, those aged 55 or older comprised 33% of smart speaker ownership, while 14- to 18-year-olds comprised only 10% [23]. Smart speakers are currently available from providers such as Amazon, Google, and Apple, and vary in shape, size, and cost, with some models having screens and cameras. For example, the Amazon Echo also contains a "drop-in" capability where devices, even if in different locations, can be linked and act as an intercom. Most also have the added capability of controlling appliances such as light bulbs, smart doorbells, and heating systems to create a "smart home." Popular uses for speakers include music, information seeking, and entertainment [24,25], but they may extend to companionship, health care support, and better sleep [26]. Speech input and output provide increased accessibility for all users, but particularly for those with limited mobility and vision.

Hampshire and Oxfordshire County councils implemented Alexa devices for 60 people requiring support from social care [27,28]. Both pilot studies found improvements in service users' ability to remain independent and feel less isolated. In Wales, the Innovate Trust investigated how smart speakers could meet the needs of adults with learning differences and reduce staff workload, thus saving money [29]. They estimated a potential saving of £20,000 (US $26,434) per year across two assisted living sites. There are an increasing number of studies examining the use of smart speakers among community-dwelling older adults [30], with qualitative feedback demonstrating commands on first interaction, which included asking for health care-related questions [31]. Although the results of a survey showed that half of care home staff think artificial intelligence in devices such as smart speakers should be in use to help care for residents [32], we found no evidence of the previous implementation of smart speakers in care homes at the majority of sites across the region under study.

The eHealth Productivity and Innovation in Cornwall and the Isles of Scilly (EPIC) project aims to develop the eHealth sector working with both the demand—improving the capability and capacity for using digital technologies—and the supply through supporting small companies producing new products and services. Early workshops identified loneliness as a major health problem and lack of skills in using technology in care homes as a barrier [33]. Voice technology was identified as an area where local expertise could be expanded, but it was clear that its use needed to be normalized [34], and a large user base would provide the incentive of a local market.

This study aimed therefore to address the loneliness and mental well-being of care home residents and at the same time to stimulate the improved uptake of technology in the care home sector, raising awareness and normalizing the use of video calls and voice-activated technologies among care home staff. We aimed to create an expectation of use across the sector.

We aimed to give at least one smart speaker device to 150 (65.2%) of 230 Cornish residential care homes and to explore if and how the devices were used, the barriers to their implementation, and their potential benefits.

Methods

Study Design

This mixed methods implementation study used care home staff diaries and telephone surveys to assess the use and impact of smart speakers. We took an eclectic approach to theory, with a lexicon borrowing ideas of "local demonstrability" [18], "technology acceptability" [16,17], implementation "at scale" [13,14], and "normalization" [15]. This eclectic conceptual framework drew mostly from NPT. The 4 NPT components were considered in our aim to understand if the devices were set up and used (cognitive participation), what they were used

for (coherence), what barriers were experienced (collective action), and any potential benefits to residents and staff (reflexive monitoring). Ethical approval was granted by the Faculty of Health Ethics Committee (reference 18/19-1054).

Choice of Smart Speakers

We aimed to implement devices that offered the possibility of video calls through a screen. In December 2018, the most appropriate device on the market was the Amazon Echo Spot (Figure 1; Multimedia Appendix 1). At Amazon's suggestion, the Amazon Kindle Fire (Multimedia Appendix 1) was trialed in some homes. Early feedback indicated that Echo Spot devices were more physically robust, and Kindle Fires looked more like "regular tablets," which some homes owned and had limited use for. With this feedback, Echo Spots were distributed until October 2019, when they were discontinued. Subsequently, we offered homes the Echo Show 5 (Multimedia Appendix 1).

Figure 1. Amazon Echo Spot.

Care Homes

Using published lists [35] updated through local knowledge, 230 homes were identified. All homes were approached at least once, using a combination of the recruitment methods.

Recruitment and Device Distribution

Initially, care homes were invited to 2 workshops, in West and East Cornwall. The participants were presented with project information and received written information, and those who agreed to participate were provided with a device. In total, 18 homes attended the 2 workshops, and 17 agreed to take part. Moreover, 4 further drop-in sessions were held over 5 months to enable care home staff to speak to researchers about the project. Those agreeing to participate were provided with a smart speaker. Drop-in sessions were advertised via postal and email campaigns and targeted telephone calling. Attendance at local drop-in sessions was low (6 people at 4 sessions from over 200 invited). However, all 6 agreed to participate. A substantial effort was also made via face-to-face recruitment with research assistants "cold calling" approximately 150 homes to explain the project and offer the opportunity to take a device. Cornwall Council also emailed all homes advertising the project resulting in 20 responses.

Support Offered to Care Homes in Using the Smart Speaker

Ongoing support was offered to care homes by two methods. First, the core EPIC team, who had made the initial approach and supplied the device. A monthly email newsletter was circulated with suggestions on different ways to use the device. Second, Digital Health Champions were recruited via undergraduate nursing and occupational therapy programs and students aged 16 or more, from 2 secondary schools. Digital Health Champions were asked to support care home staff in using smart speakers. EPIC team members supported Digital Health Champions with guidance on how to use the device and monthly group Skype sessions as a group.

Data Collection

Care homes were supplied with a short user guide and a diary to keep track of how and how often they were using the device, any barriers and how they were overcome, and any factors that enabled use of the device. Only 18 homes returned completed diaries; therefore, the remaining homes were contacted in early 2020 to take part in a short telephone survey to gain feedback on their device usage, their reason (if they were not using the device), and where it was located. We also asked for short descriptive accounts of their experiences including what the device was being used for and if they had experienced any problems. The responses were documented verbatim directly during the telephone call. Most (142, 95%) homes provided feedback either via diary or by telephone.

Data Analysis

Survey data were analyzed using descriptive statistics. Descriptive experiences of the device usage, gathered from diaries and telephone survey, were qualitatively analyzed using

thematic analysis. We followed the 6-step guide outlined by Braun and Clarke [36].

Results

Care Homes That Took 1 or More Devices

Of the 230 homes invited, 156 (68% initial uptake) homes took at least 1 device (Figure 2). Six homes returned their speaker, leaving 150 (65.2%) with devices (Figure 2). Dividing Cornwall into 5 areas (Figure 3) aggregating primary care networks, uptake was lower in East (44/79, 56%) and West Cornwall (12/23, 52%) compared with the other 3 areas (67%-78%; X^2=10, 4 *df*; *P*=.04). The 150 homes with devices comprised 92 for older people (estimated resident population 2099 [24]) and 50 for those with physical or mental health needs (resident population 1097). Moreover, 10 homes received a Kindle Fire, 141 an Echo Spot, and 5 an Echo Show 5.

Figure 2. Flowchart of recruitment, interventions, and data collection.

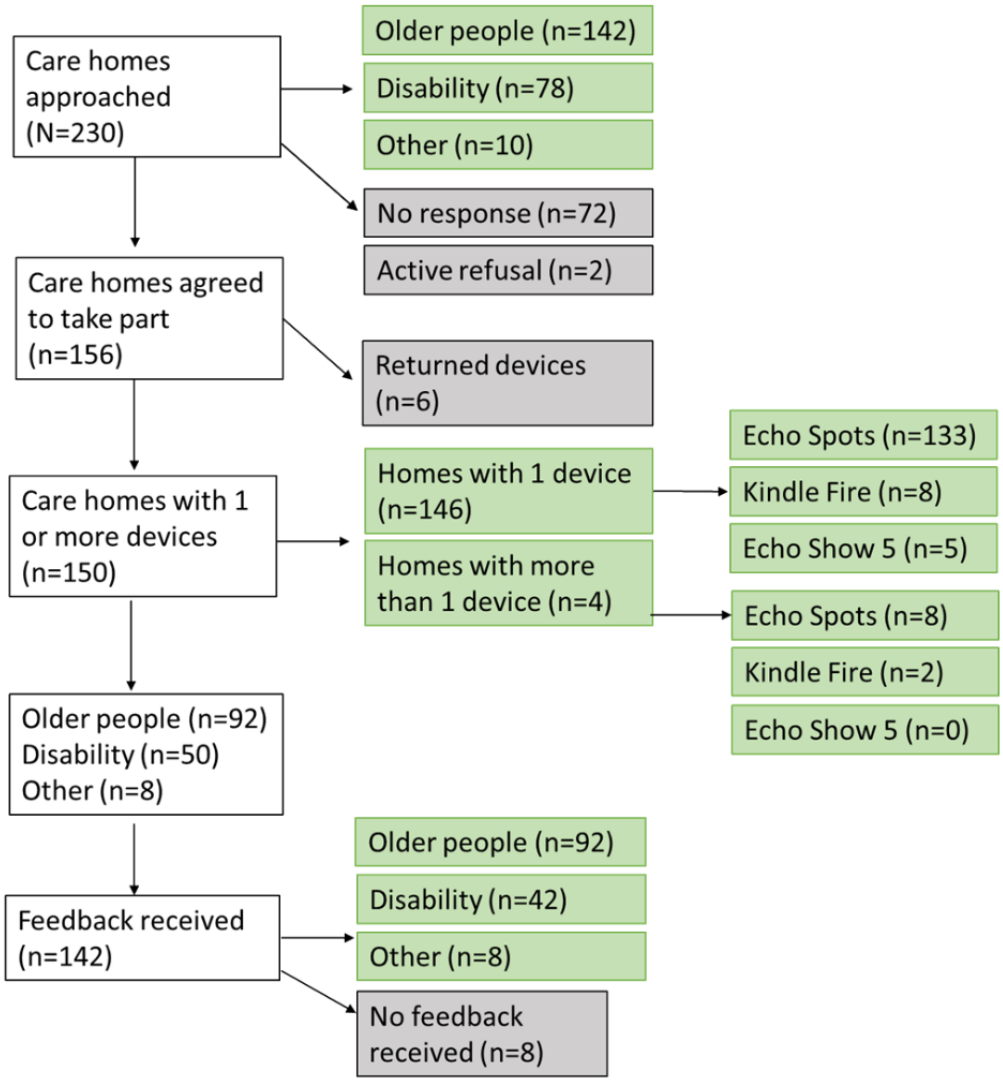
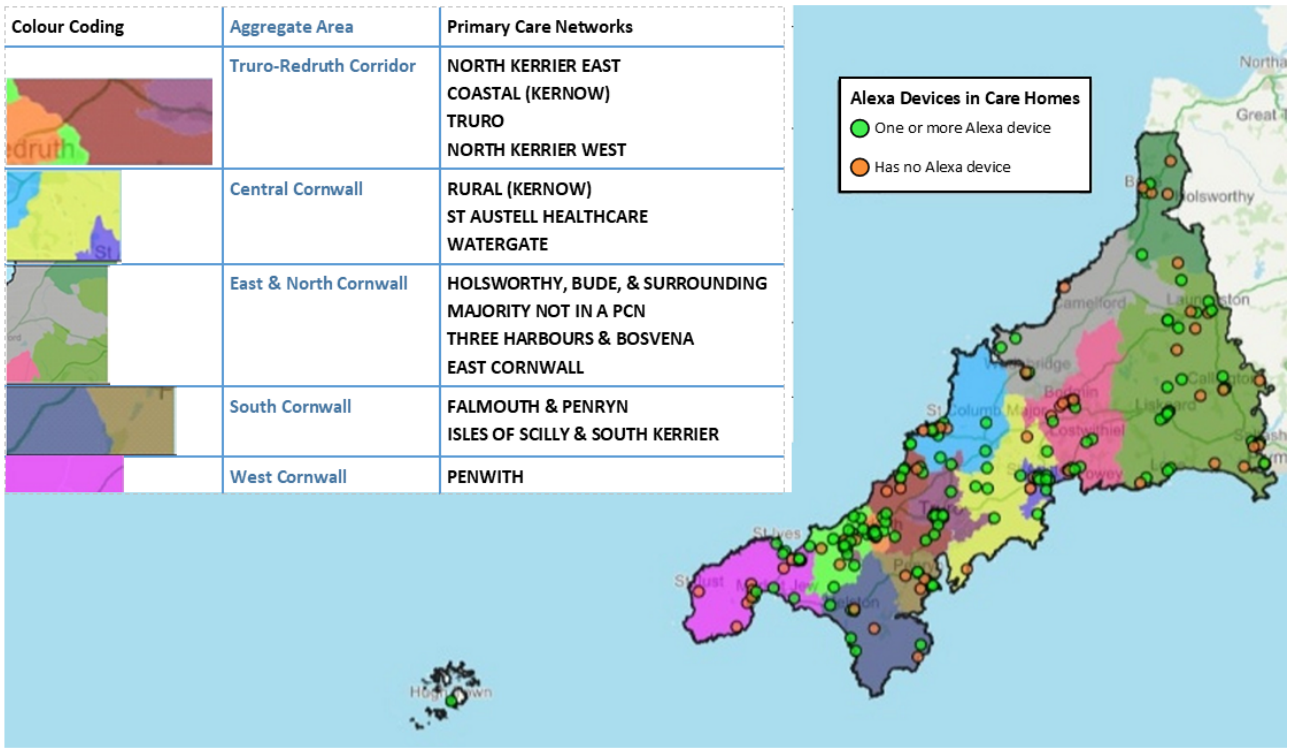


Figure 3. Location of 230 care homes in Cornwall and the Isles of Scilly showing 150 that accepted a smart speaker device and 80 that did not by primary care network areas. PCN: primary care network.



Care Homes That Returned or Did Not Take a Device

The reasons for the 6 returned devices were as follows: (1) management concerns about access to confidential information; (2) staff thinking it was inappropriate for their clients; and (3) lack of internet access. Of the 72 homes (resident population 1740), not taking devices, 3 "positively" declined but 69 did not respond.

Device Usage

Most (142, 95%) care homes provided feedback (Figure 2), of which three-quarters (107/142, 75%) were currently using the device (Table 1). More homes caring for older people were using the device than homes for people with physical and mental health needs (82% vs 61%; X^2=7.1, 2 *df*; *P*=.03). Most homes (70/107, 65%) placed the device in communal areas such as living or dining rooms. Homes caring for older people were more likely to move the device around (26/75, 35%) compared with homes for people with physical or mental health needs (2/27, 7%) (Table 1).

Table 1. Care homes using one or more devices at follow-up, by care home type.

Home type and location of use	Homes, n (%)	Homes using the device, n (%)	Homes not using the device, n (%)	Total, n
Older people		75 (82)	16 (18)	91
Moveable	26 (35)			
Communal	44 (59)			
Other	5 (7)			
Physical or mental health needs		27 (61)	17 (39)	44
Moveable	2 (7)			
Communal	22 (81)			
Other	3 (11)			
Other		5 (71)	2 (29)	7
Moveable	0 (0)			
Communal	4 (80)			
Other	1 (2)			
Total	N/A[a]	107 (75)	35 (25)	142

[a]N/A: not applicable.

Thematic Analysis

We identified 5 themes and 26 subthemes, which can be seen in Multimedia Appendix 2.

Usage

Virtually all (103/107, 96%) homes used devices for music and audio including listening to the radio, creating playlists, listening to Christmas carols, background music, and audiobooks (Multimedia Appendix 2). More than half (62/107, 58%) used it for information seeking, asking the speaker for the news, weather, and checking tide times when arranging walks. A third (35, 33%) used it for quizzes and games, 18 (17%) for jokes and comedy, and 4 (4%) for reminders. Eleven homes (10%) reported "general engagement" with residents talking to and engaging with the smart speaker. One manager reported "We have a resident who just loves to shout at Alexa, she gets in the neck but because she's compliant, the resident loves it!" Video calling was reported by only 6 (6%) homes, although some homes expressed a desire to use it but had not yet tried. However, 2 (2%) homes were using screens as part of other activities (including watching videos), 8 (7%) in sing-along sessions, and 3 (3%) using the "drop-in" feature. One (1%) found the screen useful when running quizzes or adding enjoyment to listening to music.

Moreover, 4 (4%) homes reported "advanced use" such as connecting multiple devices to create a "call bell" for one bedbound resident; "We use it as a makeshift call button, which is linked to all the other devices; when the person presses the button, a song plays across the whole house, and we know he needs support." One (1%) used the device to integrate with other smart technology.

Benefits Associated With Use

In most homes (83/107, 78%) all residents had access to the device, 10 (9%) reported 75%-90% access, and 14 (13%) with more limited access. In diaries and telephone surveys, staff relayed benefits of using the speakers despite not being directly asked. These benefits were grouped into 5 themes (Multimedia Appendix 2). The most frequently reported benefit was resident enjoyment (36/107, 34%). For example, one reported, "the person we support has started to really enjoy music and will dance, relax, or feel vibrations on the echo spots speaker," and another, "this has brought great joy and laughter to the home. Lots of dancing and singing currently taking place."

Relaxation and calming were reported by 11 (10%) homes; for example, "when a client is suffering from agitation we asked for thunderstorms and the results where amazing relaxed the client almost within minutes and he slept peacefully." Seven (7%) homes reported that the device offered companionship, one saying "Alexa is the resident's companion that never sleeps, she is always there if a resident can't sleep and wants some company." Four (4%) homes reported the acquisition of a new skill set for residents as a benefit. One said "our resident has a speech difficulty, so struggles to ask it things, but it helps her to communicate more clearly. She sees staff using it and tries to copy." The less frequently reported benefits were the time saved (2/107, 2%), increased staff engagement (3/107, 3%), and ease of use (2/107, 2%).

Barriers Associated With Nonuse

Among the 35 care homes that reported not using the device, 15 (43%) belonged to a care provider whose IT specialist raised concerns about data protection (Multimedia Appendix 2). Other reported barriers included lack of time and resource (7/35, 20%), internet connectivity (4/35, 11%), and lack of skill or confidence around technology (6/35, 17%).

Barriers Associated With Use

Most homes using their device (75/107, 70%) reported no barriers setting up and using the speaker. Homes who expressed issues most frequently reported internet connectivity problems (14/107, 13%). Nine (8%) homes reported device limitations including the smart speaker's inability to understand certain

residents because of speech difficulties or not using the "desired" language. Barriers to video calling included reluctance from family members and limitations associated with moving the device.

In 2 (2%) homes, residents had broken the device; however, both homes had repurchased devices and placed them in more secure locations. Four (4%) homes reported that the device caused confusion, or residents disliked it; however, in most cases, this was part of an adjustment period. One said, "I think originally they were a little suspicious of the speaker, but now they take it in their stride and just accept that they can ask this small 'box' questions!"

Data protection concerns were reported by 1 home with a device in use. Only 1 home reported resident inability to operate the device as a problem. A lack of skill or confidence in staff was reported by 1 home as a concern.

Support Offered and Requested

Twenty (19%) homes contacted the core EPIC team between April 2019 and February 2020. Nine (8%) homes wanted support including with setup, internet connectivity, connecting two devices, or using Skype. Four (4%) homes requested an additional device, 1 wanted a device with a bigger screen, 1 to exchange their Fire tablet with a Spot, and 5 homes requested further support from a student to use the speaker more effectively.

Moreover, 24 university and 8 secondary school students became Digital Health Champions and borrowed a device. Each university student was assigned at least 2 care homes. Of the recruited care homes, 122 (81%) were linked with a Digital Health Champions and offered support via email, phone, or occasionally face-to-face. However, only 11 homes provided the students with feedback, despite numerous attempts to contact their care homes.

Discussion

Implementation at Scale and "Normalization" of Use

We aimed to get smart speaker devices used in 65% of care homes. This proved challenging but was achieved, benefitting residents and creating a "user base" of voice-activated devices. Although at follow-up 25% (35/142) were not yet using the device, redistribution and prompting was likely to bring these devices into use. Overall, there was emerging "normalization" [34] of smart speaker use in care homes in Cornwall.

Uses and Benefits

Similar to pilot projects conducted in Oxfordshire and Hampshire [27,28], we received positive feedback on the use of smart speakers. Their use was similar to that in the general population [24,25], music and information-seeking being the most popular. While music is a basic function, its use supported opportunities for reminiscence; the speakers provided soothing music with end-of-life care, calming the residents.

Smart speakers may help reduce loneliness and increase independence [27,28]; 6 of our homes stated that the devices provided companionship for residents. Further exploration of functions that support independence, particularly for those living with physical or mental health needs, is required. For example, the "reminder" function has been shown to benefit individuals requiring support while living in their own homes [27,29,37], but only 4 of the 150 homes in our study reported spontaneous use of this function.

Similar to the findings by Pradhan et al [11], our results demonstrated unexpected uses, including 2 homes where the residents' interaction with the smart speaker provided opportunities to practice and improve expressive language skills.

Barriers to Implementation

Implementing new technology in care homes can prove challenging. Zamir et al [5] found that implementing video calls within care homes faced barriers such as staff turnover and lack of family commitment. In our study, only 6 homes reported using the devices to make video calls, suggesting that cheaper devices without screens could be used instead. However, screens were used intuitively as part of other activities including reading lyrics for sing-along sessions and displaying pictures to support information seeking. It remains to be seen how the COVID-19 pandemic has affected the use of video calls within care homes to overcome "shielding."

Studies have cited the lack of digital literacy as a barrier to smart speaker use for individuals requiring social care support at home [27]. Only 1 home in our study that used a device reported a lack of confidence or skill as a barrier; however, for those not yet using it, a lack of skill among staff was a concern. One issue experienced by homes was insufficient internet connectivity; however, outages of Wi-Fi or main power did not pose an unacceptable inconvenience, nor did they seem to inhibit continued use. Changing working methods is always difficult [29], and undertaking activities to increase digital literacy of care home staff could facilitate adoption of voice-activated technology.

As Hoy [21] also found, privacy and data protection using smart speakers was of concern for a few homes and 1 home using the device at follow-up. Such concerns therefore appeared to inhibit adoption but did not limit use once installed. Others have found patient privacy and data protection concerns where there is more specific health care use [38], yet a pilot project in Stoke-on-Trent concluded that while concerns related to privacy and accessibility cannot be ignored, smart speakers may partially solve some problems facing primary care [39]. Our research suggests that smart speaker use in communal areas of care homes for mainly entertainment purposes creates few privacy concerns. Undertaking activities to increase digital literacy including knowledge of data security of care home staff could facilitate adoption of voice-activated technology for more health-related and care-related uses.

In our study, only 1 home refused to participate based on "not liking technology." However, others may find that cultural change hinders wider adoption. Changing working methods is always difficult, as found by a Welsh project on smart speakers in supported living [29]. Undertaking activities to increase digital literacy of care home staff could facilitate adoption of voice-activated technology.

Negative Outcomes

None of the homes reported distress linked to the use of smart speakers, and only 4 homes reported temporary dislike or confusion by residents. However, confusion around the device did not typically affect enjoyment. Other reports suggest people with dementia have shown distress at having a robotic voice speaking to them [40]. How smart speakers are perceived and are of benefit to those experiencing dementia requires more in-depth exploration.

Supporting Installation of New Technology

Although all care homes had the offer of help from a Digital Health Champion student, we had few (20/150) requests for support. In some cases, Digital Health Champion students were particularly active; but overall, this aspect of the project had limited uptake as students had difficulty contacting care homes, with most requests for help being received by the core research team. This may be due to the initial personal contact made for distributing the devices.

Future Uses

As the use of smart speakers becomes normalized in care homes, one might expect an increased use for other purposes such as providing health care information or advice to staff. This extends the use beyond entertainment, and further research is needed to evaluate effective workflows [41]. Further research is also needed on designing voice user interfaces for health care, as current design training for voice interfaces is relatively limited [42].

Limitations

The feedback collected was primarily from one person, usually the manager, and therefore may only represent a single stakeholder perspective. To gain more in-depth understanding of the practical use, barriers, and facilitators of smart speakers, further research is needed from a broader stakeholder base including residents, families, and the Digital Health Champions. On collection of feedback, homes were not specifically asked to share the benefits of device usage; therefore, the benefits experienced may be understated. We had hoped that care homes would keep diaries and be able to tell us how many of their residents interacted with the smart speakers, but this data collection proved impractical. Systematic exploration of benefits and how they related to specific smart speaker "skills" and subsets of service users will be important for future research. Further exploration of the sustainability of benefits of the smart speakers is required in the longer term beyond 3-6 months.

Conclusions

This study demonstrated that most care homes are prepared to install and use smart speakers to benefit staff and residents. As an affordable and readily available commercial product, smart speakers represent a highly scalable option to facilitate technology-enabled care. Future work needs to explore how to reach the remaining care homes, deal with cybersecurity concerns, highlight beneficial "skills" for residents in the longer term, and investigate the impact on staff workload. This may lead to opportunities for smart speaker software development supported by local small and midsize enterprises.

Acknowledgments

We would like to thank Sarah Scrivener (Cornwall Council), Wo King (Hi9), Kirsty Marrs, Anita O'Connor, and other members of the EPIC project for their help and support. We would also like to thank all care homes that took part, in particular, those providing video and other feedback. Lastly, we are grateful to all University of Plymouth students who participated as Digital Health Champions and other Digital Health Champions from Scott Medical College and Pool Academy.

The hardware for this project was funded by Cornwall Council Echo Spot care home project (reference TECGS R1 8). KJE, DS, and TP were funded by the eHealth Productivity and Innovation in Cornwall and the Isles of Scilly project, a project funded in part by the European Regional Development Fund and the South West Academic Health Science Network.

Authors' Contributions

KJE recruited most care homes, distributed smart speaker devices, maintained database and data collection, called all care homes for feedback, analyzed feedback (with DS), wrote draft paper (with RBJ and DS), liaised with Cornwall Council for grant monitoring (with RBJ), and edited and approved the final paper.

RBJ had the idea for the project in response to Cornwall Council funding opportunity, carried out initial literature and informant review, successfully applied for funding and ethics permission, recruited academic colleagues to support digital health champions (with TP), bought smart speaker devices, organized initial workshops, worked with KE, DS, and TP to maintain database, carried out geographical analysis, wrote draft paper (with KJE and DS), liaised with Cornwall Council for grant monitoring (with KJE), and edited and approved the final paper.

DS worked with KJE in recruiting care homes and distributing Alexa devices, worked with KJE in gathering and analyzing feedback, wrote draft paper (with KJE and RJ), and edited and approved the final paper.

TP was coapplicant on application to Cornwall Council and University of Plymouth ethics, recruited Digital Health Champions, liaised with nursing and Occupational Therapy academics, worked with KJE, DS, and RBJ to allocate Digital Health Champions to care homes, supported Digital Health Champions (with written guides and online meetings) in trying to contact care homes, and edited and approved the final paper.

IM gave technical support in use of smart speaker devices, wrote guide and answered queries from care homes and digital health champions, and edited and approved the final paper.

TC was coapplicant on application to Cornwall Council and University of Plymouth ethics, helped with the recruitment of Digital Health Champions, and edited and approved the final paper.
HC was coapplicant on application to Cornwall Council and University of Plymouth ethics, helped with the recruitment of Digital Health Champions, and edited and approved the final paper.
AW was coapplicant on application to Cornwall Council and University of Plymouth ethics, helped recruit and manage digital health champions, and edited and approved the final paper.
FF was coapplicant on application to Cornwall Council and University of Plymouth, helped with the recruitment of Digital Health Champions, and edited and approved the final paper.
TK was coapplicant on application to Cornwall Council and University of Plymouth ethics, helped with the recruitment of Digital Health Champions, and edited and approved the final paper.
AC advised on the project and further developed links with Amazon and other stakeholders for future work using progress in this project, edited, and approved the final paper.

Conflicts of Interest

Non declared.

Multimedia Appendix 1
Amazon Echo Spot, Kindle Fire, and Echo Show 5.
[DOCX File , 1657 KB - jmir_v23i12e26767_app1.docx]

Multimedia Appendix 2
Themes, sub-themes, and example quotes.
[DOCX File , 16 KB - jmir_v23i12e26767_app2.docx]

References

1. Care homes for Older People market analysis projections. laingbuissonevents.com. UK: Laing Buisson; 2017 May. URL: https://www.laingbuissonevents.com/wp-content/uploads/2017/05/William-COP.pdf [accessed 2020-01-24]
2. Maczka M, Parry D, Curry R. Technology Innovation in Care Homes: The Sehta Review. South East Health Technologies Alliance. URL: https://www.sehta.co.uk/cms-data/depot/sehta/Technology-Innovation-in-Care-Homes-The-SEHTA-Review.pdf [accessed 2020-01-24]
3. Tingle J. The state of health care and adult social care in England. Br J Nurs 2011;20(12):760-761. [doi: 10.12968/bjon.2011.20.12.760] [Medline: 21727839]
4. SCIFE. Annual Report and Financial Statements. Social Care Institute for Excellence. 2017 Mar. URL: https://www.scie.org.uk/publications/corporate/annual-report-financial-statements-2017/annual-report-and-financial-statements-2016-17.pdf [accessed 2020-01-24]
5. Zamir S. Skype on Wheels: Implementation of video-calls to reduce feelings of lonelinesssocial isolation for older people living in care. PhD Thesis, University of Plymouth, UK. 2020. URL: https://pearl.plymouth.ac.uk/handle/10026.1/15790?show=full [accessed 2020-04-05]
6. Research and Innovation in the field of ICT for Health, Wellbeing and Ageing Well: an overview. European Commission. 2019. URL: https://ec.europa.eu/information_society/newsroom/image/document/2019-33/health_ageing_projects_list_2019_6BC92EFF-90F3-8A94-09FBFA3C4DFD150E_61321.pdf [accessed 2020-01-24]
7. Bradwell HL, Edwards KJ, Winnington R, Thill S, Jones RB. Companion robots for older people: importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England. BMJ Open 2019 Sep 26;9(9):e032468 [FREE Full text] [doi: 10.1136/bmjopen-2019-032468] [Medline: 31558461]
8. Zamir S, Hennessy CH, Taylor AH, Jones RB. Video-calls to reduce loneliness and social isolation within care environments for older people: an implementation study using collaborative action research. BMC Geriatr 2018 Mar 02;18(1):62 [FREE Full text] [doi: 10.1186/s12877-018-0746-y] [Medline: 29499659]
9. Mihailidis A, Cockburn A, Longley C, Boger J. The acceptability of home monitoring technology among community-dwelling older adults and baby boomers. Assist Technol 2008;20(1):1-12. [doi: 10.1080/10400435.2008.10131927] [Medline: 18751575]
10. Demiris G, Rantz M, Aud M, Marek K, Tyrer H, Skubic M, et al. Older adults' attitudes towards and perceptions of "smart home" technologies: a pilot study. Med Inform Internet Med 2004 Jun;29(2):87-94. [doi: 10.1080/14639230410001684387] [Medline: 15370989]
11. Pradhan A, Mehta K, Findlater L. "Accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC Canada p. 1-13. [doi: 10.1145/3173574.3174033]

12. Bentley A, King K, Radevsky A, Rainsford C. Final report: Dallas baseline economic evaluation. Databuild Research & Solutions. 2013 Mar 04. URL: https://www.winningmoves.com/file/129/download?token=89CFJqQ_ [accessed 2021-02-13]
13. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
14. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
15. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
16. Davis F. A technology acceptance model for empirically testing new end-user information systems: theory and results. MIT Libraries. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2021-02-13]
17. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
18. Rogers E. Diffusion of innovations. 5th Edition. New York, US: The Free Press; Aug 16, 2003.
19. Asthana S, Sheaff R, Jones R, Chatterjee A. eHealth technologies and the know-do gap: exploring the role of knowledge mobilisation. evid policy 2020 Nov 01;16(4):687-701 [FREE Full text] [doi: 10.1332/174426420x15808912803267]
20. Koon LM, McGlynn SA, Blocker KA, Rogers WA. Perceptions of Digital Assistants From Early Adopters Aged 55+. Ergonomics in Design 2019 Apr 25;28(1):16-23. [doi: 10.1177/1064804619842501]
21. Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Med Ref Serv Q 2018;37(1):81-88. [doi: 10.1080/02763869.2018.1404391] [Medline: 29327988]
22. Media Nations: UK 2019. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0019/160714/media-nations-2019-uk-report.pdf [accessed 2019-04-17]
23. Feldman R. Smart speaker ownership doubles in six months. YouGov. 2018 Apr 19. URL: https://yougov.co.uk/topics/politics/articles-reports/2018/04/19/smart-speaker-ownership-doubles-six-months [accessed 2020-01-24]
24. Lopatovska I, Rink K, Knight I, Raines K, Cosenza K, Williams H, et al. Talk to me: Exploring user interactions with the Amazon Alexa. Journal of Librarianship and Information Science 2018 Mar 07;51(4):984-997. [doi: 10.1177/0961000618759414]
25. Bentley F, Luvogt C, Silverman M, Wirasinghe R, White B, Lottridge D. Understanding the Long-Term Use of Smart Speaker Assistants. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2018 Sep 18 Presented at: ACM on interactive, mobile, wearable ubiquitous technologies; September 18, 2018; New York p. 1-24. [doi: 10.1145/3264901]
26. Brause SR, Blank G. Externalized domestication: smart speaker assistants, networks and domestication theory. Information, Communication & Society 2020 Jan 20;23(5):751-763. [doi: 10.1080/1369118x.2020.1713845]
27. "Alexa, Can You Support People With Care Needs?" Trialling consumer devices in adult social care. PA Consulting. 2018. URL: https://www2.paconsulting.com/rs/526-HZE-833/images/Trialling%20Consumer%20Devices%20report%20-%20PA%20Consulting.pdf [accessed 2020-01-24]
28. Hayes L. Talking Alexa: how assistive technology can help adult social care services. Community Care. 2018 Oct 18. URL: https://www.communitycare.co.uk/2018/10/18/talking-alexa-assistive-technology-can-help-adult-social-care-services/ [accessed 2020-01-24]
29. Innovate Trust: Smart Speakers for supported living. Nesta. URL: https://www.nesta.org.uk/feature/innovate-save-case-studies/innovate-trust-smart-speakers-supported-living [accessed 2020-01-24]
30. Milka T, Martin-Hammond A. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting and Abadnoning Echo. 2020 Apr 23 Presented at: CHI '20: Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3313831 [doi: 10.1145/3313831.3376760]
31. Kim S. Exploring How Older Adults Use a Smart Speaker-Based Voice Assistant in Their First Interactions: Qualitative Study. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e20427 [FREE Full text] [doi: 10.2196/20427] [Medline: 33439130]
32. Learner S. Over half of staff this care home should be using artificial intelligence to help care for residents. carehome.co.uk. 2019 Nov 25. URL: https://www.carehome.co.uk/news/article.cfm/id/1618043/quarter-care-home-staff-smart-speakers-alexa-residents [accessed 2020-01-24]
33. Jones RB, Asthana S, Walmsley A, Milligan J, Paisey M, Edwards K, et al. Developing the eHealth sector in Cornwall. Pearl, University of Plymouth. 2019 Apr. URL: https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/14904/Developing_the_eHealth_sector_in_Cornwall.pdf?isAllowed=y&sequence=1 [accessed 2020-01-24]
34. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
35. Reviews for Care Homes, Residential Homes and Nursing HOmes. carehome.co.uk. URL: https://www.carehome.co.uk [accessed 2020-04-17]

36. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. London, UK: SAGE Publications Ltd; Mar 2013.
37. Choi GS, Kim HS, Choi GH. Feasibility of smart speaker use to support aging in place. In: Innovation in aging. 2018 Nov 11 Presented at: GSA 2018 Annual Scientific Meeting; November 14-18, 2018; Boston, USA p. 560.
38. Hassoon A, Schrack J, Naiman D, Lansey D, Baig Y, Stearns V, et al. Increasing Physical Activity Amongst Overweight and Obese Cancer Survivors Using an Alexa-Based Intelligent Agent for Patient Coaching: Protocol for the Physical Activity by Technology Help (PATH) Trial. JMIR Res Protoc 2018 Feb 12;7(2):e27 [FREE Full text] [doi: 10.2196/resprot.9096] [Medline: 29434016]
39. Chambers R, Beaney P. The potential of placing a digital assistant in patients' homes. Br J Gen Pract 2019 Dec 26;70(690):8-9. [doi: 10.3399/bjgp20x707273]
40. Nicholas L, Miller C. Better Care in the Age of Automation. Doteveryone. URL: https://www.doteveryone.org.uk/wp-content/uploads/2019/09/Doteveryone_Better-Care-in-the-Age-of-Automation-1.pdf [accessed 2020-01-24]
41. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Dec 04;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
42. Murad C, Cosmin M. Designing Voice Interfaces: Back to the (Curriculum) Basics. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, USA p. A URL: https://dl.acm.org/doi/proceedings/10.1145/3313831 [doi: 10.1145/3313831.3376522]

Abbreviations

EPIC: eHealth Productivity and Innovation in Cornwall and the Isles of Scilly
NPT: Normalization Process Theory

Edited by G Eysenbach; submitted 24.12.20; peer-reviewed by B Davis, C Hennessy; comments to author 08.02.21; revised version received 17.02.21; accepted 18.11.21; published 20.12.21.

Please cite as:
Edwards KJ, Jones RB, Shenton D, Page T, Maramba I, Warren A, Fraser F, Križaj T, Coombe T, Cowls H, Chatterjee A
The Use of Smart Speakers in Care Home Residents: Implementation Study
J Med Internet Res 2021;23(12):e26767
URL: https://www.jmir.org/2021/12/e26767
doi:10.2196/26767
PMID:34932010

Original Paper

Using the Theoretical Domains Framework to Identify Barriers and Enablers to Implementing a Virtual Tertiary–Regional Telemedicine Rounding and Consultation for Kids (TRaC-K) Model: Qualitative Study

Sumedh Bele[1,2], MBBS, MPH; Christine Cassidy[3], BScN, PhD; Janet Curran[3], BScN, PhD; David W Johnson[2,4], MD; J A Michelle Bailey[2,4], MSc, MD

[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]School of Nursing, Dalhousie University, Halifax, NS, Canada

[4]Department of Pediatrics, Alberta Health Services, Calgary, AB, Canada

Corresponding Author:
J A Michelle Bailey, MSc, MD
Department of Pediatrics
Cumming School of Medicine
University of Calgary
28 Oki Dr NW
Calgary, AB, T3B 6A8
Canada
Phone: 1 403 955 3015
Email: jamichelle.bailey@albertahealthservices.ca

Abstract

Background: Inequities in access to health services are a global concern and a concern for Canadian populations living in rural areas. Rural children hospitalized at tertiary children's hospitals have higher rates of medical complexity and experience more expensive hospitalizations and more frequent readmissions. The 2 tertiary pediatric hospitals in Alberta, Canada, have already been operating above capacity, but the pediatric beds at regional hospitals are underused. Such imbalance could lead to poor patient safety and increased readmission risk at tertiary pediatric hospitals and diminish the clinical exposure of regional pediatric health care providers, erode their confidence, and compel health systems to further reduce the capacity at regional sites. A *Telemedicine Rounding and Consultation for Kids* (TRaC-K) model was proposed to enable health care providers at Alberta Children's Hospital to partner with their counterparts at Medicine Hat Regional Hospital to provide inpatient clinical care for pediatric patients who would otherwise have to travel or be transferred to the tertiary site.

Objective: The aim of this study is to identify perceived barriers and enablers to implementing the TRaC-K model.

Methods: This study was guided by the Theoretical Domains Framework (TDF) and used qualitative methods. We collected qualitative data from 42 participants from tertiary and regional hospitals through 31 semistructured interviews and 2 focus groups. These data were thematically analyzed to identify major subthemes within each TDF domain. These subthemes were further aggregated and categorized into barriers or enablers to implementing the TRaC-K model and were tabulated separately.

Results: Our study identified 31 subthemes in 14 TDF domains, ranging from administrative issues to specific clinical conditions. We were able to merge these subthemes into larger themes and categorize them into 4 barriers and 4 enablers. Our findings showed that the barriers were lack of awareness of telemedicine, skills to provide virtual clinical care, unclear processes and resources to support TRaC-K, and concerns about clear roles and responsibilities. The enablers were health care providers' motivation to provide care closer to home, supporting system resource stewardship, site and practice compatibility, and motivation to strengthen tertiary–regional relationships.

Conclusions: This systematic inquiry into the perceived barriers and enablers to the implementation of TRaC-K helped us to gain insights from various health care providers' and family members' perspectives. We will use these findings to design interventions to overcome the identified barriers and harness the enablers to encourage successful implementation of TRaC-K.

These findings will inform the implementation of telemedicine-based interventions in pediatric settings in other parts of Canada and beyond.

International Registered Report Identifier (IRRID): RR2-10.1186/s12913-018-3859-2

(J Med Internet Res 2021;23(12):e28610) doi:10.2196/28610

KEYWORDS

telemedicine; eHealth; pediatric care; inpatient; regional; rural; Canada; Theoretical Domains Framework; qualitative

Introduction

Background

The World Health Organization considers "providing equitable access to people-centered care" as one of the key components of a well-functioning health system [1]. In Canada, almost one-fifth of the population (18%) lives in rural communities, but they are served by only 8% of the physicians [2]. Evidence suggests that rural children hospitalized at tertiary children's hospitals have higher rates of medical complexity and experience more expensive hospitalizations and more frequent readmissions [3]. Although inequitable access to health services is a global concern, it is paramount for provinces in Canada to efficiently allocate health resources to reduce health care spending and provide equitable access to pediatric patients in nonmetropolitan and rural communities.

Alberta Children's Hospital (ACH) and Stollery Children's Hospital are the 2 tertiary pediatric hospitals in Alberta. Several regional hospitals throughout the province also have dedicated pediatric beds. Medicine Hat Regional Hospital (MHRH) is one such regional hospital located 300 km southeast of Calgary. Alberta Health Services (AHS) is the single health authority for Alberta province and is the largest such organization in Canada. Both ACH and MHRH are part of AHS. In 2015-16, ACH operated at 95.1% of its bed capacity half of the year and at more than 100% during the rest of the year. During the same year, MHRH showed only 57% occupancy rates. Furthermore, in that same year, 28% of the total pediatric inpatient days of stay at ACH were used by 19% of the pediatric patients from the MHRH catchment area. High bed occupancy is not only associated with poor patient safety, high mortality rates, and increased readmission risk, but it also increases the stress of health care providers [4,5]. In contrast, low use of beds in regional hospitals may diminish the clinical exposure of regional pediatric health care providers, erode their confidence, and compel health systems to further reduce the capacity at regional sites, which hampers access to specialized care for rural patients [6].

Telemedicine is a way to address tertiary–rural imbalances in health care delivery. The World Health Organization defines telemedicine as "healing at a distance" through the use of information and communication technologies (ICTs) [7]. The use of telemedicine enables health care systems to connect patients with health care providers in underserved areas, save time and travel expenses for the patients and their families, increase efficiency, and improve quality of care while supporting health care providers at rural, regional, or community health care sites [8,9]. Telemedicine is an umbrella term that incorporates the use of various forms of ICTs, including but not limited to telephone, electronic messaging, SMS text messaging, mobile apps, and audiovisual tools to provide clinical care. Among these various ICT applications, audiovisual systems are the most promising because they provide an opportunity for real-time audiovisual communication between 2 sites, which facilitates better understanding of a patient's clinical condition, more accurate diagnosis and clinical management, and enhanced communication between personnel at both sites [10]. A growing body of evidence suggests that telemedicine has the potential to bridge geographical barriers to providing clinical care, ranging from diagnosis and active management to maintaining continuity of care for pediatric patients [8]. The educational applications of telemedicine lie in its ability to connect tertiary teaching hospitals with community or rural practices, allowing bidirectional flow of information [11]. Telemedicine also enhances personal relationships between rural pediatricians and subspecialists [12].

Telemedicine-Facilitated Model for Inpatient Pediatric Clinical Care

The use of telemedicine to provide outpatient clinical care is common and usually involves a single patient visit and a few health care providers. The imbalance in use of pediatric beds in Alberta warrants testing innovative solutions to provide equitable and patient- and family-centered care to pediatric patients in nonurban and rural parts of Alberta. Therefore, *Telemedicine Rounding and Consultation for Kids* (TRaC-K), a telemedicine-facilitated model for inpatient pediatric clinical care, is in development and will be piloted between ACH and MHRH. This model was jointly developed by the team of health care providers from ACH and MHRH, as well as administrators and technical experts at AHS. TRaC-K will enable health care providers at ACH to partner with their counterparts at MHRH to provide inpatient clinical care for pediatric patients from the MHRH catchment area. The model promotes regional site access to tertiary care providers and enhances collaboration and communication between care teams. The potential patients for the TRaC-K model will be those who are either admitted at MHRH and could benefit from TRaC-K to receive care at MHRH or those from the MHRH catchment area who are admitted at ACH but are stable and could benefit from TRaC-K for potential early transfer back to MHRH to complete their remaining treatment. These patients will be identified by health care providers from both sites. A telemedicine cart with capability to provide real-time audiovisual transmission will be used for TRaC-K (Figure 1). Textbox 1 lists the key features of the TRaC-K model.

Figure 1. Mobile telemedicine cart capable of real-time audiovisual transmission.

Textbox 1. Key features of Telemedicine Rounding and Consultation for Kids (TRaC-K).

TRaC-K: key features

- Daily inpatient census and patient rounds
  - Census: Monday to Friday; health care teams review patients eligible for TRaC-K at both sites
  - Patient-focused consultation or rounds regarding the care of a patient at 1 site involving teams at both sites and may occur at the bedside with the patient and family included in care planning when able and appropriate
- Telemedicine-facilitated
  - Real-time audiovisual transmission using telemedicine cart (Figure 1)
- Tertiary–regional collaboration
  - Collaboration between Alberta Children's Hospital (tertiary) and Medicine Hat Regional Hospital (regional)
- Multidisciplinary teams
  - General pediatricians, subspecialists, allied health professionals, nurses, and nurse educators (based on patient needs)

The patient-centered–care approach of the TRaC-K model will benefit patients from regional catchment areas who will be able to access tertiary hospital–level care expertise and resources within their own region without having to travel to the tertiary-level hospital. Families will also benefit by saving travel time, minimizing time away from work, and being able to stay better connected to their community support networks.

Inpatient care at both tertiary and regional sites is provided within a complex health system with many providers and occurs independently with limited communications through telephone for specialist physician consultation related to the care of patients. Therefore, it is important to apply rigorous methods to understand the potential influences on behavior in the context in which it will occur and understand the mechanisms of change

even before implementation of the intervention. There exist several models and frameworks for assessing and identifying barriers and enablers to implementation of eHealth interventions, but their focus lies on technological aspects rather than behavioral determinants at the individual level. In the TRaC-K model, technology acts as the facilitator for the clinical model through behavior change. Implementing the TRaC-K model requires change in individual and collective behaviors of all stakeholders involved (ie, physicians, nurses, allied health providers, hospital and unit managers and administrators, patients, and families). Therefore, to systematically identify the stakeholders' perceived barriers and enablers to the implementation of the proposed TRaC-K model of inpatient pediatric care, we used the Theoretical Domains Framework (TDF). The TDF provides a theoretical lens to identify barriers and enablers to change to inform implementation of an intervention in local contexts. The TDF consists of 84 theoretical constructs that are refined into 14 domains [13,14]. The TDF has been used widely to identify behavioral determinants to implementation of health care interventions [15-17].

Therefore, a study identifying perceived barriers and enablers such as health care providers' skills, their perceptions of compatibility of the TRaC-K model, and their motivations to provide virtual care would help in the implementation of TRaC-K, especially in the local context of Canada's first and largest province-wide, fully integrated health system. A paucity of evidence in the scientific literature regarding the implementation of such interventions undermines telemedicine's potential to provide high-quality, cost-effective, patient- and family-centered, and equitable clinical care [18,19]. To our knowledge, no other study has comprehensively studied barriers and enablers to such telemedicine models of inpatient pediatric clinical care.

The COVID-19 pandemic has exponentially increased the uptake of telemedicine in clinical care; however, to our knowledge, TRaC-K is the first telemedicine-facilitated model to provide inpatient clinical care for pediatric patients. Therefore, the objective of this study is to conduct a systematic and theory-informed identification of barriers and enablers to the implementation of the TRaC-K model between ACH and MHRH.

Methods

Design

This study is part of a larger multiphase project to develop, implement, and evaluate a virtual tertiary–regional telemedicine rounding and consultation model of inpatient pediatric care. This is a qualitative descriptive study guided by the TDF. The protocol of this study has been published elsewhere [20].

Ethics

Ethical approval for this study was obtained from the Conjoint Health Research Ethics Board at the University of Calgary (REB17-1435). Administrative approval for this project was also obtained from AHS.

Setting

Participants were selected from ACH and MHRH. ACH is located in Calgary; with 141 beds, it is the largest tertiary pediatric hospital in Alberta. MHRH is located in the city of Medicine Hat, Alberta. It is a 325-bed regional hospital with 10 dedicated beds for pediatric patients, excluding the neonatal intensive care unit.

Participants

We recruited a stratified purposive sample of clinical stakeholders at ACH and MHRH. In addition, we invited family caregivers of pediatric patients with a history of availing inpatient medical services at both ACH and MHRH in the last 3 years to participate in this study. At ACH, we recruited participants by sending an email describing the study, which included an invitation to participate, to unit and allied health managers as well as physician leaders to distribute among their respective teams. We also contacted administrators, including senior leaders at ACH, through email. At MHRH, we recruited participants by sending an email describing the study, which included an invitation to participate, to site pediatric managers and the physician coinvestigator of the study who also distributed it among their respective teams. At ACH, pediatricians and Family Advisory Council members suggested the names of family caregivers meeting our criteria as potential study participants; at MHRH, a single pediatrician obtained the consent to contact from family caregivers to participate in this study. We contacted these family caregivers through email. Interested participants replied to the emails, and we contacted them to set up an in-person interview or telephone interview or focus group. Participant recruitment continued until saturation. We offered a CAD $15 (US $12) gift card honorarium to all participants for their participation in the study.

Interview Topic Guide

The TDF informed the interview guide, with 2-4 questions formulated to explore each of the 14 TDF domains (Multimedia Appendix 1). Investigators with expertise in implementation science (CC and JC) provided guidance to develop the interview guide. Next, members of the research team reviewed the guide to refine it further. We used the same guide for the focus groups, but fewer key questions were asked to initiate the discussion on different TDF domains. We modified some of the questions for the interviews with family caregivers of pediatric patients.

Procedure

A single interviewer (SB), who has expertise and experience in qualitative research, conducted all the interviews and focus groups at ACH and MHRH. For the focus groups, a notetaker accompanied the interviewer to take field notes. Before each in-person interview or focus group, we reviewed the written consent form, and each participant signed it. We excluded administrators from both the focus groups to prevent the power differential from influencing group dynamics during the focus group discussions. For the telephone interviews, we emailed the consent form to the participants before the interview and obtained written consent at the beginning of the interview. In addition, all participants from ACH and MHRH, including family caregivers, completed a brief demographic form.

Individual interviews lasted 30-60 minutes, and each focus group lasted approximately 50 minutes. We audio recorded and transcribed verbatim all the interviews and focus group discussions. Subsequently, we imported all the transcripts into NVivo software (version 11; QSR International) to code, organize, and manage the data to facilitate analysis.

Data Analysis

Before analyzing all the data, 2 reviewers from the research team (SB and CC) independently coded 2 randomly selected transcripts to discuss consistency in coding and develop a codebook. Discrepancies were resolved through discussion. On the basis of this discussion, a codebook was developed, and a single reviewer (SB) coded the remaining transcripts using this codebook.

We analyzed data in 3 steps. First, a directed content analysis approach [21] was used to categorize similar belief statements into each of the 14 TDF domains. We cross-indexed similar statements in multiple domains if they were relevant to more than one domain. Second, we used an inductive coding approach [22] to group similar belief statements to form subthemes within the initial coding scheme of the 14 TDF domains. Finally, as a study team, we further examined these subthemes to aggregate and reword before categorizing them into barriers and enablers. To prepare the results, we tabulated the subthemes within each TDF domain and tabulated larger themes into barriers and enablers separately. Quotations illustrating core beliefs are used to highlight subthemes in each domain.

Results

Interviews and Focus Groups

All the interviews and focus groups were held over a 10-month period (from November 2017 to August 2018). We conducted 31 interviews and 2 focus groups. At ACH, we conducted 15 semistructured individual interviews and 1 focus group. At MHRH, we conducted 16 semistructured individual interviews and 1 focus group. A total of 29 interviews were held face to face, whereas 2 interviews were conducted by telephone because of the inability of the participants to travel to either ACH or MHRH. The ACH focus group consisted of 5 participants, whereas the MHRH focus group included 6 participants. Both focus groups included pediatricians, nurses, and other allied health professionals. Thematic saturation where no new information was achieved was reached after interviewing 42 participants (see Table 1 for characteristics of the participants).

Table 1. Characteristics of study participants (N=42).

Characteristics and category	Participants
Gender, n (%)	
Male	3 (7)
Female	39 (93)
Site, n (%)	
Alberta Children's Hospital	20 (48)
Medicine Hat Regional Hospital	17 (40)
Family members	5 (12)
Position, n (%)	
Administrator	7 (16)
General pediatrician	8 (19)
Pediatric subspecialist	2 (5)
Nurse	10 (24)
Allied health professional	10 (24)
Family member	5 (12)
Focus group participants, n (%)	
Alberta Children's Hospital	5 (45)
Medicine Hat Regional Hospital	6 (55)

During data analysis, key statements demonstrating the beliefs of participants were attributed to each TDF domain. Next, the statements in each domain were grouped to form subthemes. Multimedia Appendix 2 lists all 14 domains of the TDF, their definition, subthemes, and representative quotes for each subtheme.

Barriers and Enablers to Implementing TRaC-K

The subthemes identified in each domain (Multimedia Appendix 2) were further categorized and tabulated into barriers and enablers. This study identified 4 barriers and 4 enablers to the implementation of the TRaC-K model, which are presented in Table 2.

Table 2. Telemedicine Rounding and Consultation for Kids (TRaC-K) barriers and enablers.

Themes and subthemes	Theoretical Domains Framework domain source
Barriers	
Awareness of telemedicine	
Limited awareness about the use of telemedicine in pediatric clinical care	1
Skills to provide virtual clinical care	
Lack of skills to communicate over the screen	2
Lack of clinical assessment skills to provide care over the screen	2
Lack of technical skills	2
Processes and resources to support TRaC-K	
Unclear processes as a potential source of harm	5, 6
Considering challenging clinical circumstances	4, 6, 10, 11
Physical environment	10
Absence of dedicated personnel	14
Difficulties in scheduling	11
Paucity of professional guidelines	3
Provider roles and responsibilities	
Concerns about clear roles and responsibilities	4
Lack of workflow integration	4, 6
Increased workload and competing priorities	11, 7
Enablers	
Motivation to provide care closer to home	
Desire to provide care closer to home	3
Confidence in TRaC-K	5
High importance	8
Excitement	13
System resource stewardship	
Balancing provincial resources	3, 6
Ability to provide tertiary-level care at regional sites	4
Redistribution of patient load and resources; care closer to homes	3, 7
Site and practice compatibility	
Compatible with current practice	9
Buy-in from key stakeholders	12
Education for potential TRaC-K users	14
Motivation to strengthen tertiary–regional relationships	
Opportunity for trust building	3, 6
Opportunity for educational exchange	3, 7

Barriers

Theme 1: Awareness of Telemedicine

Overall, there was limited awareness about the use of telemedicine in pediatric clinical care. All participants were aware of telemedicine in some form such as telephone calls or Skype video meetings among health care providers or between health care providers and patients. Most respondents had heard about telemedicine in adult care or in other jurisdictions of Canada or in the United States. Participants from MHRH were aware of outreach clinics where Medicine Hat patients have direct telemedicine consults with their pediatric specialists at ACH, but MHRH pediatric unit staff are not involved in these consults. Some of the participants reported that they have used or heard of telemedicine being used to provide specific types

of care, such as transition of care or use of synchronous audiovisual telemedicine by the patient transport team.

At the time of the interviews, awareness about using synchronous audiovisual technology for rounding and consultation in inpatient pediatric care was very low. Almost all participants reported that this was the first time they were hearing about a telemedicine-based model of inpatient pediatric care.

Theme 2: Skills to Provide Virtual Clinical Care

Participants identified a number of skills that were currently lacking and would need to be developed to use TRaC-K successfully. All respondents acknowledged the difference between in-person and virtual communication; therefore, they underlined effective communication as an important skill for the TRaC-K model of care. Family members also highlighted that lack of proper communication skills would hamper developing trust and building rapport between patients and health care providers and among health care provider teams. Health care providers at both sites reported that well-organized communication with an ability to verbalize findings, being concise with the information, and framing proper questions would be essential components of communication skills required to provide care through the TRaC-K model. MHRH providers also pointed out that as health care providers at a regional site, MHRH team members would sometimes have to be assertive during the TRaC-K rounds.

In terms of using clinical skills in a virtual health environment, health care providers at ACH believed that they would have to rely on team members at MHRH for clinical observations. In addition, they would have to improve their skills to virtually assess patients over the screen. In contrast, health care providers at MHRH believed that because of TRaC-K, they would have to handle patients with higher acuity; therefore, they might have to advance their assessment skills to make decisions about a transfer to ACH. They might also have to advance their practical skills to perform new or less familiar procedures.

Finally, participants also stressed the importance of technical skills to handle the telemedicine cart and their ability to troubleshoot any technical issues during rounds.

Theme 3: Processes and Resources to Support TRaC-K

Participants identified that lack of clear processes and resources could be an important barrier to successfully use TRaC-K. Most of the participants did not think there was any harm in providing care using the TRaC-K model, but several participants identified some potential harms such as overreliance on technology and incorrect use of this mode of providing care if processes were unclear. Health care providers mentioned a need for specific processes for using TRaC-K in challenging clinical situations as well as ensuring resources to support care in these situations.

Participants highlighted the inability to perform physical examinations and potential miscommunication as sources of adverse outcomes for patients. Health care providers raised concerns about potential delay or failure in recognizing and managing patients whose condition was deteriorating because patients' conditions may change quickly. Health care providers mentioned that because of lack of resources at MHRH, their decision to directly transfer patients to ACH instead of using TRaC-K would be influenced by the need to treat patients classified as acute, especially those with life-threatening emergencies such as respiratory arrest and cardiac arrest as well as those requiring intubation. In addition, health care providers listed clinical situations when it would be difficult to provide care using TRaC-K, such as those requiring *hands-on* care by specialists as well as those involving mental health issues, palliative care, child abuse cases, patient counseling, and patients with complex care needs. Therefore, it was suggested to establish good criteria to determine the kinds of patients who are appropriate to receive care using TRaC-K and those who are not appropriate.

Health care providers also noted several family factors that could contribute to challenging situations, such as non–English-speaking patients, cultural minorities, and families with significant stressors. They highlighted the importance of acceptance of TRaC-K by patients and their families. Participants suggested that one of the ways to mitigate these issues was to educate patients and their families about the TRaC-K model by explaining the reasons for using TRaC-K when they are admitted at MHRH or ACH and addressing any concerns they might have. Other forms of support, such as the Language Line telephone interpreting service to provide care for non–English-speaking patients and billing codes for MHRH pediatricians, were also identified by a few MHRH health care providers.

Specific resources such as the setup of the physical environment and availability of personnel for coordination of TRaC-K were mentioned by many participants. According to participants, factors related to the physical environment, such as the ease of moving the TRaC-K cart among different rooms, size of the patient rooms, and privacy in these rooms, might hinder the use of TRaC-K.

According to many participants, the implementation of TRaC-K involves several logistical issues such as testing technology, scheduling, and coordinating between the 2 sites. Scheduling TRaC-K rounds as part of the workflow and background coordination were consistently highlighted as some of the major challenges. Participants stressed the importance of having support in place before providing care using the TRaC-K model, and they indicated that having dedicated TRaC-K coordinators at ACH and MHRH would address many barriers.

All participants were asked if they were aware of any practice guidelines from their professional organizations regarding telemedicine-facilitated care. Interestingly, almost all were unsure or unaware of any relevant practice guidelines. Many of them guessed that the use of telemedicine would be encouraged by their professional organization but were not able to provide any concrete information about what type of guidelines their organizations want them to adhere to while providing telemedicine-facilitated inpatient care.

Theme 4: Provider Roles and Responsibilities

Health care providers at both ACH and MHRH stressed the importance of having clear roles and responsibilities.

Considering the potential for conflict and medicolegal issues, health care providers wanted clear guidelines on who holds the primary responsibility for the patient, especially if there is disagreement among health care providers at ACH and MHRH in the presence of families.

Health care providers from ACH also raised the issue of their lack of understanding of the capabilities of regional hospitals and the staff at MHRH. The need to integrate TRaC-K workflow into current clinical structures was also highlighted by participants. Health care providers at MHRH mentioned that pediatricians there provide care at their private clinics as well as at the inpatient unit and emergency department at MHRH; therefore, finding time to schedule their TRaC-K rounds would be challenging. Similarly, health care providers at ACH and MHRH also emphasized the importance of scheduling TRaC-K rounds in advance. Health care providers at ACH also did not want to be torn between multiple sites because this may have a negative impact on the care they provide for their own patients at ACH.

Participants were asked to imagine the potential changes that would occur with the implementation of TRaC-K. Health care providers at MHRH expressed that TRaC-K would increase their workload because they would see more patients classified as acute, and because of TRaC-K, MHRH patients transferred to ACH might be transferred back to MHRH, resulting in higher use of the pediatric unit at MHRH and increasing their workload. Participants also shared their thoughts on the situation that would arise if TRaC-K was not implemented. Without TRaC-K, the status quo would continue, resulting in continuation of current practices, including calling ACH by telephone for specialist physician consultation, patients and families traveling to ACH, and imbalance in bed occupancy.

Other competing priorities were also as seen as a potential barrier. Health care providers at MHRH mentioned that they have fewer pediatricians than ACH, and most of them have their own private clinics; therefore, they have to provide care at different locations, and accordingly sometimes other competing priorities might act as a barrier to using TRaC-K.

Additional issues such as the comfort level of health care providers at MHRH to handle patients with complex care needs and non–English-speaking patients were also mentioned as some of the circumstances under which it would be difficult to use TRaC-K. Some patient-related challenges in the use of TRaC-K were also pointed out. According to a health care provider at ACH, some patients “like” to be at ACH for various nonclinical reasons such as an opportunity to shop and visit the city.

Enablers

Theme 1: Motivation to Provide Care Closer to Home

Participants at ACH and MHRH stressed the importance of providing care closer to home, especially for pediatric patients. Participants believed that pediatric patients and their parents develop their social support mechanism in their own communities; therefore, taking patients to ACH disconnects them from such social support. The long commute (3-4 hours, one way) between MHRH and ACH and the dangers associated with such long-distance travel, especially during snowy winter months, was stated as another reason for providing care closer to home.

Most of the participants showed high confidence in TRaC-K and affirmed that it would not drastically change the current practice of providing care but rather enhance it. During the focus group at ACH, a few health care providers cautioned against overvaluing TRaC-K as a total replacement for face-to-face care. Despite several potential sources of harms having been identified, participants supported TRaC-K and expressed that the benefits of using TRaC-K would outweigh potential harms or negative consequences.

Participants also felt that the TRaC-K project was of high importance. They were asked to rate the importance of providing care through the TRaC-K model on a scale of 1-10, with 10 being very important. The score given by participants ranged from 5 to 10; most rated it as 8 or 9. The reasons for giving high scores were the TRaC-K model’s ability to keep patients and families within local communities and its potential to improve the quality of care. Some of the reasons cited for lower scores were a potential increase in workload and skepticism about the TRaC-K model’s usefulness.

In general, participants were excited about the TRaC-K model mainly because they want to try something new and TRaC-K offers them a novel way to care for their patients. However, without having tried it, some participants mentioned being optimistically cautious.

Theme 2: System Resource Stewardship

Participants acknowledged the importance of telemedicine-facilitated pediatric care in promoting sustainability from a health system’s perspective. Participants were aware of overcapacity at ACH and underuse of pediatric beds at MHRH; therefore, they recognized that the TRaC-K model would help in balancing the patient load. According to participants, TRaC-K would also facilitate transition of care for patients with complex chronic conditions from ACH to MHRH. Therefore, TRaC-K was viewed as a mechanism to address this imbalance and to advocate for additional resources for regional sites such as MHRH. None of the participants mentioned any personal monetary incentives to use the TRaC-K model of inpatient clinical care. However, the potential cost saving for AHS was mentioned as a financial incentive at the system level.

Providing care at regional sites when possible was deemed important for the system and for families. All participants reiterated the important role of TRaC-K in providing patient- and family-centered care in local communities for patients from regional sites. Other key benefits such as prompt expert care, faster diagnosis and treatment, and better quality of care were also mentioned. Health care providers acknowledged that using TRaC-K would help to provide family-centered care, and if patients and families at MHRH felt that they were not at the best hospital, then joint rounds using TRaC-K in their presence would show that ACH and MHRH were working as 1 team to provide the best care for their child. Family members mentioned the potential reduction in wait times to see a specialist as an incentive for them to use the TRaC-K model.

Participants described various incentives for them to use TRaC-K from their own perspectives. The biggest incentive for health care providers was the professional satisfaction gained by providing best-quality care closer to the patient's home.

Theme 3: Site and Practice Compatibility

Participants overwhelmingly agreed that TRaC-K would be compatible with their practice because it aligns with their motivation to provide patient- and family-centered care and technology is slowly changing the way care is provided; therefore, with a few modifications, they could easily incorporate TRaC-K into their practice or adapt their practice to use TRaC-K.

Participants highlighted that buy-in from all health care providers, especially from key stakeholders such as physicians and administrators, would be essential for successful implementation of TRaC-K. A participant even considered the interviews and focus groups for this study as an educational opportunity for them to know more about TRaC-K.

Many participants alluded to the importance of educating potential users of TRaC-K, including health care providers at ACH and MHRH and patients and their families. The educational activities mentioned were creating educational material, performing hands-on trials with technology, and practicing in mock scenarios. It was suggested that to promote family-centered care, families should be engaged right from the beginning, with explanations of the TRaC-K model provided to them verbally or by using pamphlets and of what they should expect during their stay at MHRH, instead of a screen being wheeled directly into their room and placed in front of them.

Theme 4: Motivation to Strengthen Tertiary–Regional Relationships

Participants from MHRH affirmed that having access to, and frequent interactions with, the teams from ACH would help build trust between regional and tertiary health care providers. Participants at MHRH pointed out that having access to tertiary health care providers, including specialists at ACH, through TRaC-K would enhance patients' and families' trust in MHRH and encourage them to seek care at MHRH instead of traveling to ACH, especially in situations where visiting ACH was unnecessary.

During joint rounds, clinical teams from ACH would be able to disseminate knowledge about new treatment guidelines, procedures, and protocols for handling patients with complex care needs among their peers at MHRH. ACH providers would also be able to learn about the challenges and capacities of regional sites. Participants from MHRH also viewed TRaC-K as an opportunity to learn and receive more support from their colleagues at ACH. Therefore, implementation of the TRaC-K model was considered a great learning opportunity for health care providers at both ACH and MHRH, which would enable both sides to share clinical knowledge. Regular communication between regional and tertiary health care providers would also serve as additional support for newly trained clinicians moving to regional sites because they would be able to receive second opinions and have a new set of eyes looking at their patients.

Discussion

Principal Findings

Our study identified 31 major subthemes in 14 domains, ranging from administrative issues to specific clinical conditions. These subthemes were further aggregated into major themes and categorized into 4 barriers and 4 enablers. Our findings suggest that most of the barriers concern uncertainties associated with potential hands-on, day-to-day use of TRaC-K. We anticipate that some of these barriers will be mitigated once the TRaC-K model pilot begins. On the basis of these barriers, dedicated TRaC-K coordinators will be hired to address administrative issues such as scheduling, providing technical support, and streamlining the process to reduce some of the workload. Some of the barriers identified in this study, such as lack of information technology skills and paucity of professional guidelines, are consistent with those identified in a study from Australia [12]. The ongoing COVID-19 pandemic has significantly affected the delivery of health care. Travel restrictions and the burden on tertiary sites have created a unique opportunity to use telemedicine. The urgency of providing efficient clinical care during this pandemic has mitigated many barriers to using telemedicine; however, the use of telemedicine in the postpandemic era remains uncertain. Many health care systems have realized the importance of telemedicine but still struggle to use it as a long-term solution [23]. Thus, the findings of this study remain timely and relevant.

On a positive note, the enablers demonstrate the willingness of health care providers to embrace the change to provide patient- and family-centered care. The enablers were also associated with participants' desire to change the current imbalance in health care use between ACH and MHRH and a positive attitude toward TRaC-K's potential to change the status quo and probably open new opportunities for educational exchange and trust building between tertiary and regional sites. ACH and MHRH are part of AHS, a single province-wide health authority, which might have helped participants to understand the importance of rebalancing resources between tertiary and regional sites. The potential of these enablers will be harnessed to encourage health care providers to use TRaC-K.

Next, the findings from this study will be used to map potential barriers and enablers to the behavior change techniques (BCTs) from BCT Taxonomy version 1, which is a standardized list of 93 hierarchically clustered BCTs [24]. The BCT-mapping exercise will be conducted with the key stakeholders. The BCTs that are likely to change behavior will be used as ways to mitigate barriers and harness the potential of enablers. Finally, the revised TRaC-K model will be piloted between ACH and MHRH for 1 year. The results of the TRaC-K pilot study will determine its feasibility and scale-up throughout the province.

The province-wide application of the TRaC-K model may increase regional pediatric bed use at multiple regional sites, thereby having a positive impact on the current tertiary overcapacity crisis in Alberta. In addition to addressing the imbalance in the use of pediatric capacities between regional and tertiary hospitals, it is anticipated that this project will enhance tertiary–regional collaboration, thereby supporting

numerous other provincial goals such as provincial pediatric guideline implementation and pediatric workforce sustainability. AHS has created various strategic clinical networks (SCNs) as engines for research and innovation as well as to act as vehicles for translating evidence into practice to improve patient care and health system performance. This project is supported by the Maternal, Newborn, Child & Youth SCN. Therefore, the findings of this study will be disseminated as pediatric health services research–generated knowledge through the Maternal, Newborn, Child & Youth SCN to drive the pilot and scale-up of TRaC-K in Alberta.

Interventions such as the TRaC-K model could fail to achieve the expected outcomes because of the lack of a scientific approach to identify and address factors such as barriers and enablers before implementation. This study contributes a comprehensive and systematic inquiry into perceived barriers and enablers to implementing telemedicine to the body of scientific literature. The evidence generated from this study would not only benefit other health care systems interested in implementing telemedicine-facilitated interventions to provide inpatient clinical care, but also serve as a publication to guide other research teams undertaking similar research to identify barriers and enablers for similar interventions within their own contexts.

The systematic and theoretical framework–driven approach to identify potential barriers and enablers to the TRaC-K mode is a clear strength of this study. In addition, a major strength of this study was the diversity of perspectives captured through interviews and focus groups. The inclusion of families of pediatric patients who frequently travel between ACH and other regional sites provided patient- and family-centric insights regarding the TRaC-K model. However, the findings of this study must be interpreted keeping some limitations in mind. First, this study was conducted before the COVID-19 pandemic; therefore, some of the barriers and enablers might have changed. Methodologically, the TDF is a framework; hence, it only describes empirical phenomena by classifying them into sets of domains. The TDF neither describes nor provides an explanation for a phenomenon; nor does it specify a relationship among its domains. Although we excluded administrators from the focus groups to mitigate the potential power imbalance, pre-existing professional relationships might have influenced the opinions of some of the participants of these focus groups. Finally, recruitment through stratified sampling required contacting family caregivers meeting the eligibility criteria through their pediatricians, which might have resulted in selection bias.

Conclusions

This systematic inquiry into perceived barriers and enablers to the implementation of TRaC-K helped us to gain insights from health care providers' and family members' perspectives. We are optimistic that the implementation of TRaC-K will be successful based on the enablers identified through this study. Enablers such as motivation to provide care closer to pediatric patients' homes and to adjust the imbalance in health care resource use will play a key role in the implementation of TRaC-K. The association of barriers with the uncertainties concerning day-to-day use of TRaC-K will enable us to address these barriers by creating clear processes and providing support through dedicated staff. Finally, these findings will inform the development and implementation of telemedicine-based interventions in other parts of Canada and beyond.

Acknowledgments

This study was made possible by support from the Maternal, Newborn, Child & Youth Strategic Clinical Network (MNCY SCN). The authors would like to gratefully acknowledge the study participants (health care providers and patients) for their contributions to this study. The authors would also like to thank Mr Ashton Chugh for helping with notetaking during the focus groups.

The project team acknowledges funding from the Department of Pediatrics, University of Calgary, and the Health Outcomes Improvement Fund from MNCY SCN. SB received financial support in the form of a graduate studentship from MNCY SCN, Collaborative Research and Innovation Opportunities scholarship from the Ward of the 21st Century Research and Innovation Center at the University of Calgary, and Graduate Student Scholarship from the Alberta Children's Hospital Research Institute.

Authors' Contributions

JAMB, DWJ, JC, and CC contributed to the overall study rationale, design, and methods development. SB conducted all the interviews and focus groups. SB and CC analyzed the data. SB led the drafting of the manuscript. All authors contributed to the drafting of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for health care providers.
[DOCX File , 27 KB - jmir_v23i12e28610_app1.docx]

Multimedia Appendix 2
Subthemes identified in each domain of the Theoretical Domains Framework.

[DOCX File , 20 KB - jmir_v23i12e28610_app2.docx]

References

1. Key components of a well functioning health system. World Health Organization. 2010. URL: http://bibalex.org/baifa/en/resources/document/450818 [accessed 2021-11-26]
2. Wilson CR, Rourke J, Oandasan IF, Bosco C. Progress made on access to rural healthcare in Canada. Can J Rural Med 2020;25(1):14-19 [FREE Full text] [doi: 10.4103/CJRM.CJRM_84_19] [Medline: 31854338]
3. Commodari E. Children staying in hospital: a research on psychological stress of caregivers. Ital J Pediatr 2010 May 25;36:40 [FREE Full text] [doi: 10.1186/1824-7288-36-40] [Medline: 20500854]
4. Madsen F, Ladelund S, Linneberg A. High levels of bed occupancy associated with increased inpatient and thirty-day hospital mortality in Denmark. Health Aff (Millwood) 2014 Jul;33(7):1236-1244. [doi: 10.1377/hlthaff.2013.1303] [Medline: 25006151]
5. Friebel R, Fisher R, Deeny SR, Gardner T, Molloy A, Steventon A. The implications of high bed occupancy rates on readmission rates in England: a longitudinal study. Health Policy 2019 Aug;123(8):765-772 [FREE Full text] [doi: 10.1016/j.healthpol.2019.06.006] [Medline: 31262535]
6. Aragona E. Ongoing trends in pediatric hospitalizations. In: Proceedings of the AAP Experience National Conference and Exhibition. 2015 Presented at: AAP Experience National Conference and Exhibition; Oct 24-27, 2015; Washington, DC URL: https://aap.confex.com/aap/2015/webprogram/Paper30119.html
7. Telemedicine: opportunities and developments in member states. World Health Organization. 2010. URL: http://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2021-11-26]
8. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
9. Marcin JP, Ellis J, Mawis R, Nagrampa E, Nesbitt TS, Dimand RJ. Using telemedicine to provide pediatric subspecialty care to children with special health care needs in an underserved rural community. Pediatrics 2004 Jan;113(1 Pt 1):1-6. [doi: 10.1542/peds.113.1.1] [Medline: 14702439]
10. Waller M, Stotler C. Telemedicine: a Primer. Curr Allergy Asthma Rep 2018 Aug 25;18(10):54. [doi: 10.1007/s11882-018-0808-4] [Medline: 30145709]
11. Olson CA, McSwain SD, Curfman AL, Chuo J. The current pediatric telehealth landscape. Pediatrics 2018 Mar;141(3):e20172334. [doi: 10.1542/peds.2017-2334] [Medline: 29487164]
12. Ray KN, Demirci JR, Bogen DL, Mehrotra A, Miller E. Optimizing telehealth strategies for subspecialty care: recommendations from rural pediatricians. Telemed J E Health 2015 Aug;21(8):622-629 [FREE Full text] [doi: 10.1089/tmj.2014.0186] [Medline: 25919585]
13. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
14. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
15. Debono D, Taylor N, Lipworth W, Greenfield D, Travaglia J, Black D, et al. Applying the Theoretical Domains Framework to identify barriers and targeted interventions to enhance nurses' use of electronic medication management systems in two Australian hospitals. Implement Sci 2017 Mar 27;12(1):42 [FREE Full text] [doi: 10.1186/s13012-017-0572-1] [Medline: 28347319]
16. Lawton R, Heyhoe J, Louch G, Ingleson E, Glidewell L, Willis TA, ASPIRE programme. Using the Theoretical Domains Framework (TDF) to understand adherence to multiple evidence-based indicators in primary care: a qualitative study. Implement Sci 2016 Aug 08;11:113 [FREE Full text] [doi: 10.1186/s13012-016-0479-2] [Medline: 27502590]
17. Thomas S, Mackintosh S. Use of the theoretical domains framework to develop an intervention to improve physical therapist management of the risk of falls after discharge. Phys Ther 2014 Nov;94(11):1660-1675. [doi: 10.2522/ptj.20130412] [Medline: 25035266]
18. Olson CA, Thomas JF. Telehealth: no longer an idea for the future. Adv Pediatr 2017 Aug;64(1):347-370. [doi: 10.1016/j.yapd.2017.03.009] [Medline: 28688597]
19. Marcin JP, Shaikh U, Steinhorn RH. Addressing health disparities in rural communities using telehealth. Pediatr Res 2016 Jan;79(1-2):169-176. [doi: 10.1038/pr.2015.192] [Medline: 26466080]
20. Bele S, Cassidy C, Curran J, Johnson DW, Saunders C, Bailey JM. Barriers and enablers to implementing a virtual tertiary-regional Telemedicine Rounding and Consultation (TRAC) model of inpatient pediatric care using the Theoretical Domains Framework (TDF) approach: a study protocol. BMC Health Serv Res 2019 Jan 11;19(1):29 [FREE Full text] [doi: 10.1186/s12913-018-3859-2] [Medline: 30634969]
21. 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]

22. Williamson K, Given LM, Scifleet P. Qualitative data analysis. In: Research Methods (Second Edition). Cambridge, UK: Chandos Publishing; 2018:417-439.
23. 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]
24. 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]

Abbreviations

ACH: Alberta Children's Hospital
AHS: Alberta Health Services
BCT: behavior change technique
ICT: information and communication technology
MHRH: Medicine Hat Regional Hospital
MNCY SCN: Maternal, Newborn, Child & Youth Strategic Clinical Network
SCN: strategic clinical network
TDF: Theoretical Domains Framework
TRaC-K: Telemedicine Rounding and Consultation for Kids

Edited by R Kukafka; submitted 08.03.21; peer-reviewed by B Nievas Soriano, KL Mauco; comments to author 10.05.21; revised version received 27.05.21; accepted 04.11.21; published 22.12.21.

Please cite as:
Bele S, Cassidy C, Curran J, Johnson DW, Bailey JAM
Using the Theoretical Domains Framework to Identify Barriers and Enablers to Implementing a Virtual Tertiary–Regional Telemedicine Rounding and Consultation for Kids (TRaC-K) Model: Qualitative Study
J Med Internet Res 2021;23(12):e28610
URL: https://www.jmir.org/2021/12/e28610
doi:10.2196/28610
PMID:34941561

Original Paper

The Use of Multimode Data Collection in Random Digit Dialing Cell Phone Surveys for Young Adults: Feasibility Study

Daniel Alexander Gundersen[1], PhD; Jonathan Wivagg[2], PhD; William J Young[3], PhD; Ting Yan[2], PhD; Cristine D Delnevo[3,4], MPH, PhD

[1]Survey and Qualitative Methods Core, Division of Population Sciences, Dana-Farber Cancer Institute, Boston, MA, United States

[2]Westat, Rockville, MD, United States

[3]Center for Tobacco Studies, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

[4]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

Corresponding Author:
Daniel Alexander Gundersen, PhD
Survey and Qualitative Methods Core
Division of Population Sciences
Dana-Farber Cancer Institute
450 Brookline Avenue, LW 686
Boston, MA, 02215
United States
Phone: 1 6176325827
Email: DanielA_Gundersen@dfci.harvard.edu

Abstract

Background: Young adults' early adoption of new cell phone technologies have created challenges to survey recruitment but offer opportunities to combine random digit dialing (RDD) sampling with web mode data collection. The National Young Adult Health Survey was designed to test the feasibility of this methodology.

Objective: In this study, we compared response rates across the telephone mode and web mode, assessed sample representativeness, examined design effects (DEFFs), and compared cigarette smoking prevalence to a gold standard national survey.

Methods: We conducted a survey experiment where the sampling frame was randomized to single-mode telephone interviews, telephone-to-web sequential mixed mode, and single-mode web survey. A total of 831 respondents aged 18 to 34 years were recruited via RDD at baseline. A soft launch was conducted prior to main launch. We compared the web mode to the telephone modes (ie, single-mode and mixed mode) at wave 1 based on the American Association for Public Opinion Research response rate 3 for screening and extended surveys. Base-weighted demographic distributions were compared to the American Community Survey. The sample was calibrated to the US Census Bureau's American Community Survey to calculate DEFFs and to compare cigarette smoking prevalence to the National Health Interview Survey. Prevalence estimates are estimated with sampling weights and are presented with unweighted sample sizes. Consistency of estimates was judged by 95% CI.

Results: The American Association for Public Opinion Research response rate 3 was higher in the telephone mode than in the web mode (24% and 30% vs 6.1% and 12.5%, for soft launch and main launch, respectively), which was reflected in response rate 3 for screening and extended surveys. During the soft launch, the extended survey and eligibility rate were low for respondents pushed to the web mode. To boost productivity and survey completes for the web condition, the main launch used cell phone numbers from the sampling frame where the sample vendor matched the number to auxiliary data, which suggested that the number likely belonged to an adult in the target age range. This increased the eligibility rate, but the screener response rate was lower. Compared to population distribution from the US Census Bureau, the telephone mode overrepresented men (57.1% [unweighted n=412] vs 50.9%) and those enrolled in college (40.3% [unweighted n=269] vs 23.8%); it also underrepresented those with a Bachelor of Arts or Science (34.4% [unweighted n=239] vs 55%). The web mode overrepresented White, non-Latinos (70.7% [unweighted n=90] vs 54.4%) and those with some college education (30.4% [unweighted n=40] vs 7.6%); it also underrepresented Latinos (13.6% [unweighted n=20] vs 20.7%) and those with a high school or General Education Development diploma (15.3% [unweighted n=20] vs 29.3%). The DEFF measure was 1.28 (subpopulation range 0.96-1.93). The National Young Adult Health Survey cigarette smoking prevalence was consistent with the National Health Interview Survey overall (15%, CI 12.4%-18% [unweighted 149/831] vs 13.5%, CI 12.3%-14.7% [unweighted 823/5552]), with notable deviation among

18- to 24-year-olds (15.6%, CI 11.3%-22.2% [unweighted 51/337] vs 8.7%, CI 7.1%-10.6% [unweighted 167/1647]), and those with education levels lower than Bachelor of Arts or Science (24%, CI 19.3%-29.4% [unweighted 123/524] vs 17.1%, CI 15.6%-18.7% [unweighted 690/3493]).

Conclusions: RDD sampling for a web survey is not feasible for young adults due to its low response rate. However, combining this methodology with RDD telephone surveys may have a great potential for including media and collecting autophotographic data in population surveys.

(J Med Internet Res 2021;23(12):e31545) doi:10.2196/31545

KEYWORDS

web mode; web survey; random digit dialing; mixed mode surveys; survey methodology; data capture; research methods; recruitment; survey; feasibility; smoking

Introduction

For a multitude of reasons, young adults have been a difficult and resource-demanding population on whom to conduct surveillance research using traditional sampling and data collection methods, especially for longitudinal surveys. They are highly mobile, moving at over twice the rate of other adults [1], making address-based and area-based sampling, as well as recontacting for follow-up data collection, challenging and resource-demanding. Likewise, they are much more likely than the general adult population to live in nontraditional and group housing such as military barracks or college dormitories [2]. Moreover, they are early adopters of new communication technologies that replace traditional contact modes, such as wireless substitution (eg, foregoing traditional landline telephones in favor of only owning cell phones). The combination of early adoption of wireless substitution, nontraditional housing, and high mobility has created substantial challenges to conducting both repeated cross-sectional and longitudinal telephone-based surveillance research among this important population subgroup. For example, their high and rapid rate of wireless substitution from the early 2000s through today led to a decline in participation and likely increased coverage bias in traditional random digit dialing (RDD) surveillance systems such as the Behavioral Risk Factor Surveillance System [3,4]. Likewise, their higher rate of living in institutional housing leads to greater coverage errors for many surveillance studies that use address-based or area probability sampling. Lastly, their high mobility rate means they are more difficult to locate and contact for follow-up data collection in longitudinal surveillance designs that typically rely on numerous approaches to engage participants such as mailings, keeping contact information for someone who knows the respondent (in case of difficulty recontacting), and making telephone calls. This is particularly detrimental to rapid surveillance of behaviors and health conditions that are themselves highly dynamic or occur in dynamic environments, such as tobacco use, which has seen a drastic shift in the tobacco product market with the introduction and growth of emerging products such as e-cigarettes in a historically short period of time.

Paradoxically, wireless substitution proved beneficial with respect to sampling young adults for cross-sectional telephone surveillance surveys [5]. Indeed, cell phone ownership is practically universal among young adults, and research found that cell-phone–only RDD led to a representative cross-section of the young adult population [3,6]. However, despite advances in sampling and weighting methodologies to address the wireless substitution challenges, response rates have continued to drastically decline. For example, Pew Research Center reports an average response rate of 28% in 2001 compared with only 6% in 2018 [7]. Thus, considerable challenges remain in conducting behavioral surveillance research, particularly rapid surveillance, in this population. This recent drastic decline in response rates for RDD surveys may be related to the rapid transformation of cell phone technology and, in turn, how cell phones are used, particularly among young adults who are the early adopters of this technology [8,9]. Indeed, smartphones, which have seen an incredibly rapid uptake in the population, allow users to communicate via multiple modes—voice, text, email, messaging, and through social media. People are decreasingly using the device for telephone conversations and increasingly using them to connect to the internet for communication via social media, texting, and email [9,10]. Moreover, some cell phone platforms (eg, iOS, Apple Inc) now include functions that block incoming calls that are not from telephone numbers known to the recipient [11]. Thus, RDD surveys once again are faced with major challenges related to secular shifts in technology use. However, as with wireless substitution, these challenges may create a new opportunity by providing a single device through which young adults can be sampled, recruited, and administered surveys via multiple modes for both cross-sectional and longitudinal surveillance research projects that use RDD and telephone mode. Indeed, this could save tremendous resources and lay the groundwork for the collection of new types of data if smartphones can be used for administering web surveys with data quality that is on par with telephone surveys.

We designed the National Young Adult Health Survey (NYAHS), a national longitudinal RDD survey of young adults, to compare two ways of integrating web-based data collection—sequential telephone-to-web mixed mode and single-mode web survey—against traditional single-mode telephone surveys. This paper uses the baseline data collection of the NYAHS to compare the response rates across telephone and web modes; to assess the generalizability of the sample by benchmarking the demographic distributions to population distributions; to examine the impact of sampling design and sample weighting on statistical precision; and to compare estimated cigarette smoking prevalence to the prevalence from a gold standard national survey.

Methods

Sampling and Recruitment

The NYAHS was designed to provide representative cross-sectional and longitudinal estimates of health behaviors using RDD sampling and to test the feasibility and utility of integrating web surveys in surveillance systems that use RDD sampling. Specifically, we compared three modes of data collection: (1) single-mode telephone survey, (2) sequential telephone-to-web dual mode, and (3) single-mode web survey. Single-mode telephone survey is the traditional approach for RDD surveillance systems, while sequential telephone-to-web dual mode and single-mode web surveys are experimental. The sample was purchased from Marketing Systems Group, and the initial sample of 136,000 cell phone numbers was randomized to one of the three data collection modes prior to loading to the computer-assisted telephone interviewing system. Figure 1 presents the sampling and original data collection design for the study. For each mode, a random selection of cellular phone numbers from cellular-dedicated thousand-level blocks were selected. Telephone interviewers manually dialed selected numbers and read a recruitment script that explained the purpose of the study and screened for eligibility. For single-mode telephone, recruitment and each data collection occasion were conducted entirely through telephone. For the sequential mixed mode, the baseline or wave 1 survey as well as waves 2 and 3 were conducted using telephone recruitment and data collection (identical to single-mode telephone), while brief follow-up surveys were conducted via the web. For the single-mode web survey, baseline recruitment was conducted via telephone with the web mode data collection. Following baseline, invitations to web surveys were sent via text and/or email depending on the respondents' preference.

Figure 1. NYAHS sampling and original data collection design. CATI: computer-assisted telephone interview; NYAHS: National Young Adult Health Survey; RDD: random digit dialing.

Frame		Mode	Sampling + recruitment mode	Wave 1 mode	Brief follow-up	Wave 2 mode	Brief follow-up	Wave 3 mode
Telephone numbers	Randomization	Single-mode telephone	RDD + CATI	CATI	CATI	CATI	CATI	CATI
		Sequential mixed mode	RDD + CATI	CATI	Web	CATI	Web	CATI
		Single-mode web	RDD + CATI	Web	Web	Web	Web	Web

To improve retention, a revision was made to the design to collect data via telephone for all modes following wave 2.

We worked 28,519 samples as a soft launch to identify ways to streamline data collection. To ensure adequate sample size for tobacco types with lower prevalence, "ever use" of any tobacco product (ie, having used a tobacco product at any time, hereafter "ever tobacco") was an inclusion criterion. However, in combination with lower response rates in the web condition, the ever tobacco inclusion criterion proved infeasible. Thus, for the main survey launch, the only inclusion criterion was being 18 to 34 years old. Moreover, in order to increase the likelihood that we reach eligible people in the single-mode web condition, we only used telephone numbers that were identified as having a high likelihood of belonging to young adults based on an additional preprocessing step using auxiliary data. We also made small wording changes to the recruitment script. No other changes were made to the recruitment process or data collection for either of the conditions.

Instrumentation

The survey instrument was designed to collect data to estimate annual tobacco prevalence and longitudinal tobacco use trajectories. Consequently, most items in the survey asked about the use of tobacco and nicotine products, but the survey also asked about the awareness of new and emerging tobacco products, cannabis and alcohol use, and several demographic characteristics. We used standardized measures when possible and adapted nonstandardized measures from other tobacco surveys. The English language survey was translated into Spanish by a bilingual researcher and independently reviewed by a second translator with inconsistencies adjudicated by the investigator team. New survey questions were pretested in Spanish and English using cognitive interviewing techniques, and a usability test was carried out on iOS and Android (Google Inc) devices.

Data Collection

Baseline data collection occurred between April 2018 and May 2019 via a computer-assisted telephone interviewing system and web-based survey. A screening questionnaire was used to identify eligible participants, defined as young adults between the ages of 18 and 34 years. The respondents in the single-mode telephone and sequential telephone-to-web modes were immediately transitioned to the main interview following the recruitment script and eligibility determination, while the respondents in the single-mode web group were texted a link

to a URL for the web survey. The URL led them to a consent form page after which the respondents were taken to the main survey. Interviews and web surveys were conducted in Spanish when language barriers were encountered during recruitment. The interviews took an average of 16.8 minutes to complete and were comparable in length in both English and Spanish. Invitation texts and emails were sent in Spanish for Spanish-speaking respondents. The participants were offered a conditional US $15 electronic gift card to a major online retailer as an incentive. The data collection protocol was approved by the Institutional Review Board at Rutgers Biomedical Health Sciences.

Sampling Weights

A base weight was calculated as the product of the inverse of the probability of selection for each sample member and the number of cellular telephones on which a respondent receives calls. The base weight was adjusted for a 24% oversampling of ever tobacco users due to the inclusion criterion of the soft launch. The base weight was then calibrated via iterative proportional fitting to population numbers from the 2018 American Community Survey (ACS) [12], a large-scale survey conducted by the Census Bureau to update population estimates between decennial censuses. Specifically, the base-weighted sample was calibrated along age (18-21 years, 24-29 years, and 30-34 years), sex (male and female), education (<BA vs ≥BA), and race or ethnicity (White, non-Latino; Black, non-Latino; Asian, non-Latino; Latino; and other). Population demographic distributions were generated from weighted analysis of the 664,617 participants aged 18 to 34 years in the ACS Public Use Microdata Sample. Calibration of the NYAHS was conducted using the 'survey' package in R (R Core Team) [13,14].

Statistical Analysis

The eligibility rate was calculated as the number of respondents meeting the inclusion criteria (aged 18-34 years and ever using tobacco during the soft launch; for the main launch, just ages 18-34 years) divided by the total number of respondents who were contacted and who answered the screening questions about age (and tobacco use early in the study). Second, we calculated the American Association for Public Opinion Research response rate 3 for screening, for extended surveys, and overall [15]. The screening response rate 3 is calculated as the number of completed screeners (eligibility questions) divided by the estimated number of sampled phone numbers active and used for personal calls. Importantly, it includes in the denominator an estimate of the eligible numbers among the dialed numbers where eligibility could not be determined. Since a vast number of the dialed numbers fall in this category, it is instructive to look at the overall response rates as well as the extended survey (ie, among those for whom eligibility could be established), the latter of which is also known as cooperation rate.

Sample representativeness across data collection modes was assessed by comparing the base-weighted NYAHS demographic distribution for sex (male and female), age (18-21 years, 22-24 years, 25-29 years, and 30-34 years), and race or ethnicity (White, non-Latino; Black, non-Latino; Asian, non-Latino; Latino; and other) against population distribution from the ACS. Single-mode telephone and sequential telephone-to-web mixed mode was collapsed as their protocols are identical at baseline. Because the calibration of base-weighted demographic distributions to population values is based only on point estimates, we did not present confidence intervals as a measure of precision. Rather, we used the National Center for Health Statistics Data Presentation Standards for Proportions criteria for reliability to suppress data that are not reliable [16].

The impact of sampling design and calibration on statistical precision is summarized with the sampling design effect (DEFF). DEFF is the ratio of variance under the complex sampling design to the variance under a simple random sample for a given estimate. We calculated DEFF as presented by Kish [17], who incorporates the correlation of sampling weights with the outcome, and presented it overall and by key subpopulations.

To assess bias in cigarette smoking prevalence, we compared the estimated prevalence from the NYAHS combined across data collection modes to the National Health Interview Survey (NHIS) [18]. The NHIS is an in-person interviewer-mediated national survey that uses area probability sampling. Thus, it is not subject to any bias resulting from an RDD approach or due to push-to-web survey following telephone recruitment. Cigarette smoking was selected as a comparison benchmark as its measurement has long been standardized. By contrast, newer products do not have uniform question wording across surveillance systems; thus, they lack a gold standard measurement. Current smokers were defined as individuals who have smoked at least 100 cigarettes in their lifetime and who currently smoke every day or some days. We presented point estimates with 95% CI calculated using the logit method. The NHIS and NYAHS estimates were generated using final sampling weights and are presented with unweighted numerators and denominators. Consistency of prevalence estimates were judged by comparing the point estimates and the degree of overlap of confidence intervals. The estimates were suppressed from tables and figures if they did not meet the National Center for Health Statistics Data Presentation Standards for Proportions [16]. All analyses were conducted using the survey package in R [13,14].

Results

Eligibility and response rates are presented in Table 1, for the main data collection (ie, after including the supplemental sampling frame for the single-mode web survey and implementing minor tweaks to the introductory language) and the soft launch. The eligibility rate for telephone was 12% (143/1191) for the soft launch and 17.6% (864/4909) for the main data collection, and 9.6% (51/532) and 30.3% (306/1009), respectively, for the single-mode web survey (where only cell phone numbers identified as having a higher likelihood of being in the 18-34 years age range were used). The screening response rate was 32.8% (4909/14980) and 39.8% (1191/2994) for the main and soft launch, respectively, compared to 16.1% (1009/6248) and 42.6% (532/1248) for the single-mode web survey. The extended response rate was 68.6% (593/864) and 74.8% (107/143) in the main and soft launch for the telephone survey and 37.9% (116/306) and 29.4% (15/51), respectively, for the single-mode web survey. These resulted in the overall

response rates of 24% (main) and 30% (soft launch) for telephone and 6.1% (main) and 12.5% (soft launch) for the single-mode web survey.

Table 1. Response rate by mode from the National Young Adult Health Survey.

Eligibility and response rate	Soft launch			Main		
	Phone	Mixed mode	Web	Phone	Mixed mode	Web
Eligibility rate, n/N (%)	71/588 (12.1)	72/603 (11.9)	51/532 (9.6)	448/2393 (18.7)	416/2516 (16.5)	306/1009 (30.3)
Screener response rate 3, n/N (%)	588/1496 (39.3)	603/1498 (40.3)	532/1248 (42.6)	2393/7763 (30.8)	2516/7218 (34.8)	1009/6248 (16.1)
Extended response rate 3, n/N (%)	55/71 (77.5)	52/72 (72.2)	15/51 (29.4)	306/448 (68.3)	287/416 (69)	116/306 (37.9)
Response rate 3, %	30.4	29.1	12.5	21.1	24	6.1

The base-weighted sample demographics for telephone and single-mode web survey benchmarked against the 2018 ACS are presented in Table 2. The telephone survey tracked the overall population distribution well by race or ethnicity and age. The NYAHS telephone survey slightly overrepresented men (57.1% [unweighted n=412] vs 50.9%) and those currently enrolled in college (40.3% [unweighted n=269] vs 23.8%) and underrepresented those with a Bachelor of Arts or Science (34.4% [unweighted n=239] vs 55%). The single-mode web sample closely tracked the population based on gender, but overrepresented White, non-Latinos (70.7% [unweighted n=90] vs 54.4%) and those with some college education (30.4% [unweighted n=40] vs 7.6%) and underrepresented Latinos (13.6% [unweighted n=20] vs 20.7%), and those with a high school or General Education Development diploma (15.3% [unweighted n=20] vs 29.3%).

Table 2. National Young Adult Health Survey demographic distribution (base weighted) benchmarked to the 2018 American Community Survey for 18- to 34-year-olds.

Demographics and category	Census (ACS[a])	Phone or mixed (unweighted N=700)	Web (unweighted N=131)
Gender, unweighted n (weighted %)			
Female	49.1	288 (42.9)	58 (45.4)
Male	50.9	412 (57.1)	73 (54.6)
Age (years), unweighted n (weighted %)			
18-24	40.4	309 (45.5)	28 (22.0)
25-29	30.6	188 (26.4)	49 (37.3)
30-34	29.0	203 (28.1)	54 (40.7)
Race or ethnicity, unweighted n (weighted %)			
Asian, non-Latino	6.5	55 (8.6)	N/A[b]
Black, non-Latino	14.3	80 (11.8)	11 (8.0)
Latino	20.7	148 (21.2)	20 (13.6)
Other, non-Latino	4.1	44 (6.2)	N/A
White, non-Latino	54.4	373 (52.2)	90 (70.7)
Education, unweighted n (weighted %)			
<HS[c]	8.2	58 (8.1)	N/A
HS or GED[d]	29.3	146 (20.9)	20 (15.3)
Some college	7.6	257 (36.7)	40 (30.4)
BA[e] or BS[f]	55.0	239 (34.4)	68 (51.8)
Currently enrolled in college, unweighted n (weighted %)			
Yes	23.8	269 (40.3)	39 (30.8)
No	76.2	431 (59.7)	92 (69.2)

[a]ACS: American Community Survey.

[b]N/A: not applicable (does not meet National Center for Health Statistics Data Presentation Standards for Proportions).

[c]HS: high school.

[d]GED: General Educational Development.

[e]BA: Bachelor of Arts.

[f]BS: Bachelor of Science.

Table 3 displays the DEFF overall and by subpopulations. Overall, the estimated variance of the sampling distribution was 1.28 times greater than if the same estimate had been derived from a simple random sample. This corresponds to an approximately 13% inflation of the estimated standard errors compared with simple random sampling. DEFF ranged from a low of 0.96 among those with a Bachelor of Arts or Science to 1.93 among those with Some College, corresponding to a range of 2% decrease to 39% increase in standard errors compared to simple random samples, respectively.

Table 3. Design effect by subpopulations.

Demographics	DEFF[a]
Gender	
Female	1.29
Male	1.38
Age (years)	
18 to 24	1.60
25 to 29	1.25
30 to 34	1.27
Race or ethnicity	
Asian, non-Latino	1.29
Black, non-Latino	1.50
Latino	1.25
Other, non-Latino	1.08
White, non-Latino	1.28
Education	
<HS[b]	1.11
HS or GED[c]	1.10
Some college	1.93
BA[d] or BS[e]	0.96
Overall	1.28

[a]DEFF: design effect.

[b]HS: high school.

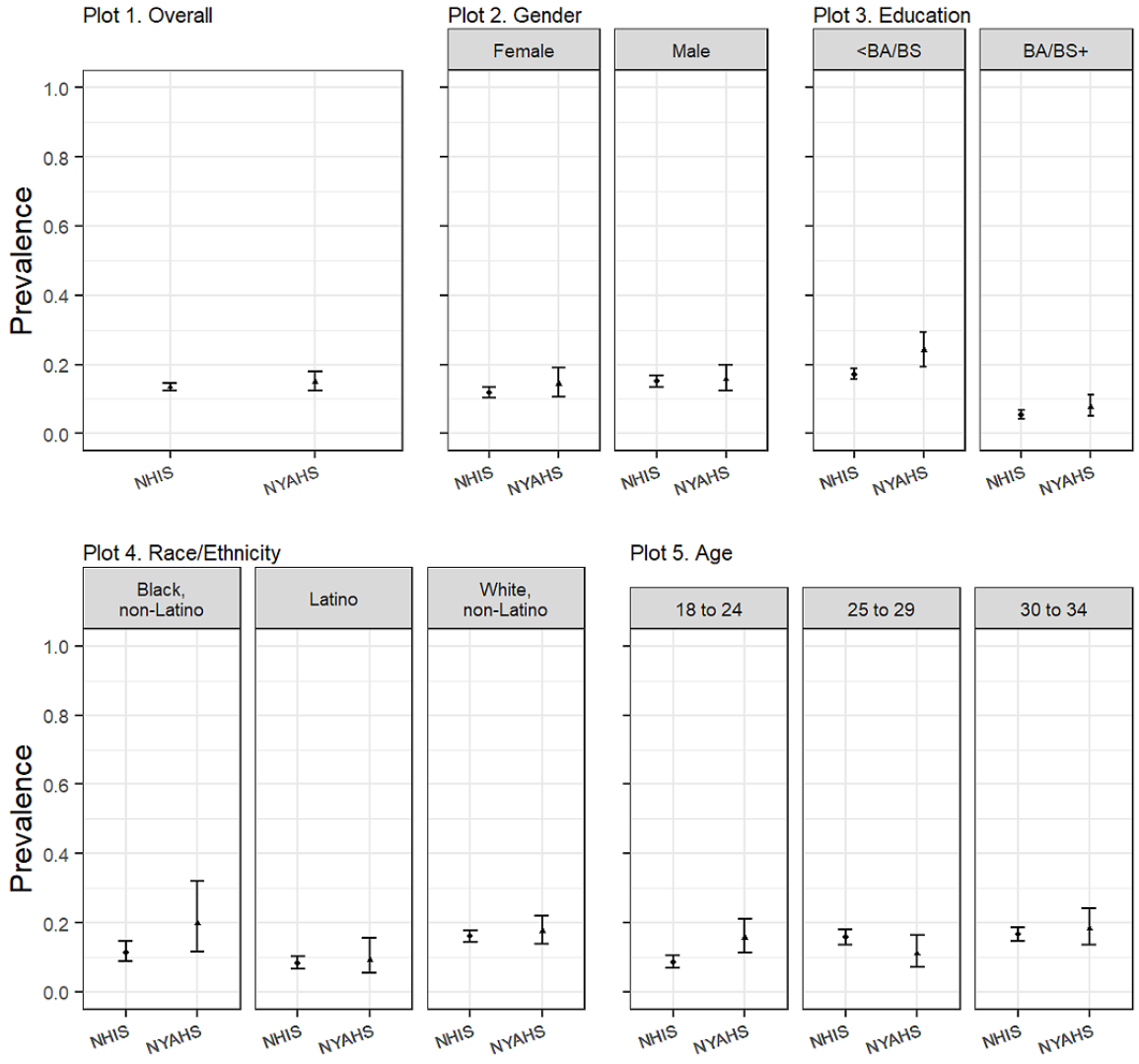
[c]GED: General Education Development.

[d]BA: Bachelor of Arts.

[e]BS: Bachelor of Science.

Figure 2 shows the estimated smoking prevalence for the NYAHS and NHIS overall (plot 1) and population subgroups (plots 2 through 5). The estimated prevalence was consistent overall (NYAHS: 15%, 95% CI 12.4%-18% [unweighted 149/831]; NHIS: 13.5%, 95% CI 12.3%-14.7% [unweighted 823/5552]), and for men, women, Whites, Latinos, Black, non-Latinos, 30- to 34-year-olds, among those with education levels <Bachelor of Arts or Science, and those with Bachelor of Arts or Science. There were notable differences only among 18- to 24-year-olds (NYAHS: 15.6%, CI 11.3%-22.2% [unweighted 51/337]; NHIS: 8.7%, CI 7.1%-10.6% [unweighted 167/1647]), and those with education levels <Bachelor of Arts or Science (NYAHS: 24%, CI 19.3%-29.4% [unweighted 123/524]; NHIS: 17.1%, CI 15.6%-18.7% [unweighted 690/3494]).

Figure 2. Cigarette smoking prevalence, NYAHS vs NHIS. BA: Bachelor of Arts; BS: Bachelor of Science; NHIS: National Health Interview Survey (2018); NYAHS: National Young Adult Health Survey (2018).



Estimates for Asian, non-Latino, and Other, non-Latino populations do not meet National Center for Health Statistics Data Presentation Standards for Proportions.

Discussion

Principal Findings

Our analysis demonstrated that using RDD sampling and telephone recruitment to a single-mode web survey is not feasible as a sole methodology for surveillance research, but it suggests that web surveys may have good utility using concurrent mixed mode designs (ie, concurrently conducting telephone and web surveys with different respondents) for surveillance research that use RDD sampling and telephone data collection mode. Our soft launch found that eligibility rates, including the use of tobacco as a qualifier, were so low that it proved cost prohibitive using traditional RDD sampling frames. Removing the requirement for tobacco use improved the recruitment using RDD sampling, though low response rates and the limited age eligibility range still made this a resource-intensive effort. In particular, response rates to the web survey among respondents screened as eligible were very low, though they were similar to those reported by Pew [7], which is consistent with other push-to-web research projects [19]. To increase web survey completes, the main baseline launch used an additional preprocessing step of the sampling frame to increase the likelihood of reaching a young adult. Indeed, this nearly tripled the eligibility rate.

Our methodology had good representativeness for the telephone mode and, to a lesser extent, the web mode. For the web condition, the base-weighted sampling distribution tracked the population distribution well on most characteristics, except for overrepresenting older young adults and White, non-Latinos. By contrast, the telephone sample tracked the population very well across all demographics, except college enrollment. This suggests that the preprocessed list-assisted RDD sampling frame we used for the web mode may have greater coverage error and/or differential nonresponse compared with the telephone

mode. However, deviation from population distributions is common in surveys with probability sampling designs and can be addressed via weighting adjustments such as poststratification or, as in our case, calibration methods [20]. Indeed, the pooled web and telephone surveys produced unbiased cigarette estimates following calibration. Notable exceptions were for 18- to 24-year-olds and those with less than a bachelor's degree, though this may be due to differences in the NHIS's poststratification and our sample calibration approaches. In particular, the NHIS does not poststratify on college enrollment, which is known to be associated with tobacco use [21], while the NYAHS does [22]. Moreover, our approach had DEFFs that were considerably smaller than the Behavioral Risk Factor Surveillance System's national estimates of 4.49 (although that estimate varies from year to year) [23]. This ensured reasonable precision for overall and some subpopulation estimates. However, larger sample sizes may be necessary for producing estimates with adequate statistical precision for smaller subgroups, particularly with distributions that deviate substantially from the population as larger weighting adjustments may result in increased variability of sampling weights and smaller effective sample sizes. Because the web mode had poorer representativeness, sample calibration and thus precision may be impacted by the proportion of the sample that is allocated to the web mode in concurrent mixed mode surveillance designs. In our case, approximately 16% (n=131) of the total sample came from the web survey, but future research should be carried out to identify the optimal allocation of the web mode and telephone mode in order to maximize the statistical precision relative to cost. The integration of a web mode in RDD methodology has tremendous potential to collect data not possible in traditional telephone surveys and/or conducting mobile-based research that is difficult to implement using probability-based sampling designs. For example, including web surveys in a telephone-based surveillance methodology allows researchers to embed visual and auditory media [24,25]. This can be used to test social marketing messaging on a large and geographically diverse scale, which can help inform public health workers and policy makers about the effective approaches to health promotion messaging. Similarly, it allows respondents to submit media in autophotography data collection to capture important contextual or exposure information that may not be adequately captured by closed or open-ended questions [26]. Such methodologies have traditionally been used in qualitative research [27,28], but future research should explore ways in which combining web surveys with RDD sampling may be used to collect such data in large-scale surveys that use probability sampling methods. Similarly, our methodology suggests that RDD sampling and the web mode may be combined with passive data collection from health tracking data linked to the respondents' mobile devices, which may reduce the overall data collection burden on them. To date, however, the feasibility and challenges of linking survey responses to health trackers in general population surveys are not known. Future research should evaluate the receptivity of respondents to participate in pulse surveys, link the resulting information with tracking data, and identify challenges and ways to address them.

Limitations

The limitations of this paper are that representativeness was evaluated over a relatively limited number of demographic characteristics. Ours were chosen because they are the most commonly used in poststratification or calibration and are priority groups for tobacco control. Moreover, consistency of prevalence estimates should be evaluated for other health indicators. This can be challenging because many tobacco products, particularly new and emerging tobacco products, do not have standardized measurements across surveillance systems. Thus, we would not be able to disentangle differences due to measurement versus sampling and data collection mode.

Conclusions

Our findings demonstrated that integrating the web mode into the traditional telephone mode in a concurrent mixed mode design for surveillance research is challenging, and RDD sampling for web-based surveillance methodology may not be feasible as a sole data collection methodology. However, our findings also suggested that the web mode can be integrated in an RDD telephone surveillance system by allocating a random subsample to complete the same survey on the web. This has tremendous potential to enhance data collection, particularly for testing social marketing messages and combining general population surveys with autophotographic or health tracking data.

Acknowledgments

This research was supported by the National Cancer Institute grant 2R01CA149705.

Authors' Contributions

DAG, CDD, and JW contributed to the conceptual design, data analysis, and interpretation of data. WJY and TY contributed to the data analysis and interpretation of data. All authors contributed to the manuscript revision submission approval.

Conflicts of Interest

None declared.

References

1. Benetsky MJ, Burd CA, Rapino MA. Young Adult Migration 2007-2009 to 2010-2012. American Community Survey Reports. 2015. URL: https://www.census.gov/content/dam/Census/library/publications/2015/acs/acs-31.pdf [accessed 2021-06-23]
2. Total Population in Households and Group Quarters by Sex and Selected Age Groups, for the United States: 2000. United States Census Bureau. 2001. URL: https://www2.census.gov/programs-surveys/decennial/2000/phc/phc-t-07/tab01.pdf [accessed 2021-06-18]
3. Blumberg SJ, Luke JV. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June 2020. National Center for Health Statistics. 2021 Feb. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless202102-508.pdf [accessed 2021-05-19]
4. Delnevo CD, Gundersen DA, Hagman BT. Declining estimated prevalence of alcohol drinking and smoking among young adults nationally: artifacts of sample undercoverage? Am J Epidemiol 2008 Jan 01;167(1):15-19 [FREE Full text] [doi: 10.1093/aje/kwm313] [Medline: 17977896]
5. Gundersen DA, ZuWallack RS, Dayton J, Echeverría SE, Delnevo CD. Assessing the feasibility and sample quality of a national random-digit dialing cellular phone survey of young adults. Am J Epidemiol 2014 Jan 01;179(1):39-47 [FREE Full text] [doi: 10.1093/aje/kwt226] [Medline: 24100957]
6. Gundersen DA, Peters KR, Conner A, Dayton J, Delnevo CD. Stability of sample quality for a national random-digit dialing cellular phone survey of young adults. Am J Epidemiol 2014 Nov 15;180(10):1047-1049 [FREE Full text] [doi: 10.1093/aje/kwu271] [Medline: 25352574]
7. Kennedy C, Hartig H. Response rates in telephone surveys have resumed their decline. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/ [accessed 2021-06-23]
8. Mobile Fact Sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-06-23]
9. Anderson M. Mobile Technology and Home Broadband 2019. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2019/06/PI_2019.06.13_Mobile-Technology-and-Home-Broadband_FINAL2.pdf [accessed 2021-06-23]
10. Cellphone Activities 2013. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_Cell-Phone-Activities-May-2013.pdf [accessed 2021-06-23]
11. Avoid unwanted calls on iPhone. Apple. URL: https://support.apple.com/guide/iphone/avoid-unwanted-calls-iphe4b3f7823/ios [accessed 2021-06-09]
12. 2018 American Community Survey 1-year Public Use Microdata Samples. United States Census Bureau. URL: https://www2.census.gov/programs-surveys/acs/data/pums/2018/1-Year/ [accessed 2021-06-09]
13. Lumley T. Analysis of Complex Survey Samples. J. Stat. Soft 2004;9(8):1-19. [doi: 10.18637/jss.v009.i08]
14. The R Project for Statistical Computing. The R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2021-06-24]
15. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. The American Association for Public Opinion Research. 2016. URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2021-06-24]
16. Parker JD, Talih M, Malec DJ, Beresovsky V, Carroll M, Gonzalez JF, et al. National Center for Health Statistics Data Presentation Standards for Proportions. Vital Health Stat 2 2017 Aug(175):1-22 [FREE Full text] [Medline: 30248016]
17. Kish L. Weighting for unequal Pi. Journal of Official Statistics 1992;8(2):183-200.
18. National Health Interview Survey, 2018 Data Release. Centers for Disease Control and Prevention.: National Center for Health Statistics URL: https://www.cdc.gov/nchs/nhis/nhis_2018_data_release.htm [accessed 2021-06-09]
19. Lynn P. Evaluating push-to-web methodology for mixed-mode surveys using address-based samples. Survey Research Methods 2020;14(1):19-30. [doi: 10.18148/srm/2020.v14i1.7591]
20. Kolenikov S. Post-stratification or a non-response adjustment? Surv Pract 2016 Aug 01;9(3):1-12. [doi: 10.29115/SP-2016-0014]
21. Green MP, McCausland KL, Xiao H, Duke JC, Vallone DM, Healton CG. A closer look at smoking among young adults: where tobacco control should focus its attention. Am J Public Health 2007 Aug;97(8):1427-1433. [doi: 10.2105/AJPH.2006.103945] [Medline: 17600242]
22. National Health Interview Survey, 2017. Centers for Disease Control and Prevention.: National Center for Health Statistics URL: https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm [accessed 2021-06-09]
23. Iachan R, Pierannunzi C, Healey K, Greenlund KJ, Town M. National weighting of data from the Behavioral Risk Factor Surveillance System (BRFSS). BMC Med Res Methodol 2016 Nov 15;16(1):155 [FREE Full text] [doi: 10.1186/s12874-016-0255-7] [Medline: 27842500]
24. Couper MP. Technology Trends in Survey Data Collection. Social Science Computer Review 2016 Aug 19;23(4):486-501 [FREE Full text] [doi: 10.1177/0894439305278972]
25. Couper MP, Tourangeau R, Kenyon K. Picture This!: Exploring Visual Effects in Web Surveys. Public Opinion Quarterly 2004 Jun 01;68(2):255-266. [doi: 10.1093/poq/nfh013]

26. Mobile technologies for conducting, augmenting and potentially replacing surveys. American Association of Public Opinion Research. 2014. URL: https://www.aapor.org/getattachment/Education-Resources/Reports/REVISED_Mobile_Technology_Report_Final_revised10June14.pdf.aspx [accessed 2021-06-23]
27. Glaw X, Inder K, Kable A, Hazelton M. Visual Methodologies in Qualitative Research. International Journal of Qualitative Methods 2017 Dec 19;16(1):1-8. [doi: 10.1177/1609406917748215]
28. Noland C. Auto-photography as research practice: Identity and self-esteem research. Journal of research Practice 2006;2(1):1-19.

Abbreviations

ACS: American Community Survey
DEFF: design effect
NHIS: National Health Interview Survey
NYAHS: National Young Adult Health Survey
RDD: random digit dialing

Edited by R Kukafka; submitted 24.06.21; peer-reviewed by R Watson, T Hughes; comments to author 10.07.21; revised version received 15.09.21; accepted 22.09.21; published 20.12.21.

Please cite as:
Gundersen DA, Wivagg J, Young WJ, Yan T, Delnevo CD
The Use of Multimode Data Collection in Random Digit Dialing Cell Phone Surveys for Young Adults: Feasibility Study
J Med Internet Res 2021;23(12):e31545
URL: https://www.jmir.org/2021/12/e31545
doi:10.2196/31545
PMID:34932017

Original Paper

Assessing Physicians' Recall Bias of Work Hours With a Mobile App: Interview and App-Recorded Data Comparison

Hsiao-Han Wang[1,2*], MD; Yu-Hsuan Lin[1,2,3,4*], MD, PhD

[1]Department of Psychiatry, College of Medicine, National Taiwan University, Taipei, Taiwan

[2]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan

[3]Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan

[4]Institute of Health Behaviors and Community Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

[*]all authors contributed equally

Corresponding Author:
Yu-Hsuan Lin, MD, PhD
Institute of Population Health Sciences
National Health Research Institutes
35 Keyan Road
Zhunan Township
Miaoli County, 35053
Taiwan
Phone: 886 37 206 166 ext 36383
Email: yuhsuanlin@nhri.edu.tw

Abstract

Background: Previous studies have shown inconsistencies in the accuracy of self-reported work hours. However, accurate documentation of work hours is fundamental for the formation of labor policies. Strict work-hour policies decrease medical errors, improve patient safety, and promote physicians' well-being.

Objective: The aim of this study was to estimate physicians' recall bias of work hours with a mobile app, and to examine the association between the recall bias and physicians' work hours.

Methods: We quantified recall bias by calculating the differences between the app-recorded and self-reported work hours of the previous week and the penultimate week. We recruited 18 physicians to install the "Staff Hours" app, which automatically recorded GPS-defined work hours for 2 months, contributing 1068 person-days. We examined the association between work hours and two recall bias indicators: (1) the difference between self-reported and app-recorded work hours and (2) the percentage of days for which work hours were not precisely recalled during interviews.

Results: App-recorded work hours highly correlated with self-reported counterparts (r=0.86-0.88, P<.001). Self-reported work hours were consistently significantly lower than app-recorded hours by –8.97 (SD 8.60) hours and –6.48 (SD 8.29) hours for the previous week and the penultimate week, respectively (both P<.001). The difference for the previous week was significantly correlated with work hours in the previous week (r=–0.410, P=.01), whereas the correlation of the difference with the hours in the penultimate week was not significant (r=–0.119, P=.48). The percentage of hours not recalled (38.6%) was significantly higher for the penultimate week (38.6%) than for the first week (16.0%), and the former was significantly correlated with work hours of the penultimate week (r=0.489, P=.002)

Conclusions: Our study identified the existence of recall bias of work hours, the extent to which the recall was biased, and the influence of work hours on recall bias.

(J Med Internet Res 2021;23(12):e26763) doi:10.2196/26763

KEYWORDS

smartphone; mobile app; work hours; recall bias; time perception; physicians; labor policy

Introduction

Excessive work hours adversely affect physicians' alertness and performance [1], increase the number of medical errors, and jeopardize patient safety [2]. The crucial effects of work-hour policies on patient safety have been widely described since 2003 when the Accreditation Council for Graduate Medical Education (ACGME) adopted restrictions on physician work hours to 80 hours a week and no more than 24 consecutive hours [2-6]. Due to similar concerns, in 2019, the work-hour restrictions in Taiwan became stricter, from a limit of 88 hours per week adopted in 2013 to 80 hours per week and 28 hours of continuous work duty.

ACGME's regulations shifted following evidence provided by trials and systematic reviews. In 2003, the ACGME allowed 24 hours of continuous work duty, which was reduced to 16 hours in 2011 and then reverted back to 24 hours in 2017 following results of a randomized controlled trial published in 2016 showing that more flexible duty-hour policies resulted in noninferior patient outcomes and physicians' self-reported well-being when compared with restrictive policies [4]. In addition, a recent study showed that residents on 16-hour or less schedules made more serious medical errors than those working shifts spanning 24 hours or more [7].

The controversial effects of eliminating extended-duration work shifts for physicians on patient safety might result from the methodological limitations of the measurement of work hours. Previous studies of work policies focused disproportionately on consecutive work hours of night shifts rather than on total work hours within a time frame, such as work hours per week. These studies were mostly limited by using work-hour measurements yielded from self-reported or medical staff–recorded logs [7] and described the fluctuating, inconsecutive nature of physician work hours with consecutive work hours. Self-reported work hours is a widely used metric in most research [3,8,9], although it is continuously shown to be unsuitable for monitoring over longitudinal time periods. Moreover, the value of self-reported work hours might be reduced by biases, especially in the estimation of total work hours. In addition, programs' compliance with ACGME regulations are usually based on medical residents' self-reports, which might be prone to the residents' biases [10]. Resident physicians, of various proportions, have admitted to not reporting duty hours accurately so as to appear compliant with regulations [6,9].

Currently, the widespread use and deep reach of smartphones in modern life enable the measurement of work hours in an affordable, reliable, and unobtrusive way. We developed an app, "Staff Hours," to automatically calculate a user's work hours via GPS background data [11]. Staff Hours is a region-restricted app, which could only be downloaded in Taiwan. Staff Hours records consecutive work hours in real time with accuracy, and enables comparisons of work hours among different hospitals, departments, and divisions with aggregate data collected from the Staff Hours database.

Using Staff Hours, we assessed physicians' estimation of work hours and found a clear bias toward underestimation. The specific aims of this study were to (1) identify recall bias indicators, including the percentage of days that work hours were not precisely recalled during interviews (NR) and the difference between self-reported and app-recorded work hours (D); and (2) examine the correlation between these two recall bias indicators and app-recorded work hours. We hypothesized that the two recall bias indicators can effectively demonstrate how work hours influence recall bias differently during the previous week and the penultimate week before assessment.

Methods

GPS-Defined Work Hours Recorded by the App

Staff Hours is a newly designed app that captures the work hours and patterns of medical staff in real time [11]. Participants installed the Staff Hours app onto their smartphones from Apple's App Store or Android's Google Play store. This app collects objective GPS location data continuously in the background and has a power-saving design. Using geofencing technology, the app automatically records the work hours one spends in the workplace. The sensitivity (94.6%) and specificity (93.9%) of app-recorded work hours were validated in our previous study [11]. The app is illustrated in Figure 1A. For example, the user's regular work hours (6:30 AM to 5:30 PM) and on-call duty (5:30 PM to 12:00 AM) were 11 hours and 6.5 hours, respectively, on November 1; the sum of the work hours, 17.5 hours, was calculated by the app. The lower part of the screenshot reveals the total work hours of each day of the previous week. The upper part of the screenshot shows the regular work hours (66 hours) and overtime work hours (19.4 hours) on the left side, and the total work hours (85.2 hours) of the past 7 days at the top-right corner of the screen. App-generated data provide real-time work hours recorded with high temporal resolution (within 10 minutes). The app saves all recorded GPS data in a log file and uploads these data to the database every day. Figure 1B shows an example of recording the daily work hours from November 1 to December 1. This user typically had longer work hours during on-call duties (night shift), including November 1-2, 7-8, 21-22, and 24-25, as well as November 30 to December 1, on which days the scheduled works hours were 16 hours (8:00 AM to 12:00 AM). There were 10 holidays within this period of time, denoted with red figures on the x-axis. On five holidays (November 3, 9, 10, 16, 23), the user was free from on-call duty, and the work hours were 0. App-recorded work hours of the previous week (WH_{APP-1}), the week before that (WH_{APP-2}), and the previous month (WH_{APP-M}, denoted by average work hours per week in 1 month), as well as the mean (SD) work hours of regular work days within a month, were obtained from app-record data. The app-recorded work hours were used as a putative gold standard to validate self-reported work hours.

Figure 1. Screenshot of the Staff Hours app user interface, with app-recorded and self-reported work hours. (A) Weekly summary of daily work hours. The light gray bar represents regular work and the dark gray bar indicates on-call duty (night shift). (B) The Staff Hours app continuously collects GPS data without active data entry by the smartphone users and provides the daily work hours. (C) Recall bias was quantified by calculating the differences between the app-recorded (orange line) and self-reported (blue line) work hours of the last week (November 25 to December 1) and the penultimate week (November 18 to 24).

Participants

A total of 18 medical doctors (11 men; median age 29 years, range 24-53 years) were recruited by randomly selecting resident physicians working at hospitals in Taipei from August 2019 to December 2019. Approval was obtained from the ethics committee in Taiwan (approval number: EC1070107-E). The inclusion criteria were resident physicians who could install the Staff Hours app on their mobile phones. There were no strict exclusion criteria. Each participant installed the app and ran it for at least 2 months. The study duration was 61 days, contributing 1098 person-days. Some of the participants failed to record their work hours for technical reasons. When this happened, their data were considered missing data. There were 30 missing days. Hence, the final data included 1098–30=1068 person-days. The participants received identical structured interviews at the first and second month after installation of the Staff Hours app. The majority (16/18, 89%) of participants were resident physicians undertaking training programs in specialized fields, including surgery, internal medicine, pediatrics, obstetrics and gynecology, emergency medicine, occupational and environmental medicine, and psychiatry. All participants were given detailed descriptions regarding the study, and individual informed consent was obtained in written form. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Institutional Review Board of the National Health Research Institutes.

Self-Reported Work Hours

The two authors, who are psychiatrists experienced in physicians' work-hour patterns and structured interviews, conducted structured interviews at the first and second month after the app installation blinded to the app-recorded work hours. An identical structured interview was repeated 1 month after the first interview. The interrater reliability of $WH_{self\text{-}1}$, $WH_{self\text{-}2}$, NR_1, and NR_2 between the two interviewers was 1.00 in all cases based on Pearson correlation analysis. The lengths of both interviews were documented. The interviews simulated investigations conducted by the Taiwan Ministry of Labor. The definitions of work hours and nonwork hours (eg, continuing medical education) were in line with the standard policies in hospitals in Taiwan [12]. The first interview assessed recalled work hours of the past month. In this case, the interviewers recorded the average time of arriving and leaving work on regular work days over the past month, as well as how many work days and on-call days, respectively, there had been. The total work hours of the previous month were calculated by the obtained report. The second interview assessed the recalled work hours of the previous week and the penultimate week. In this case, the interviewers recorded the specific work hours of each day (from Monday to Sunday) during the previous week and the penultimate week, as demonstrated in Figure 1C. Recall bias was quantified by calculating the differences between the app-recorded (orange line) and self-reported (blue line) work hours of the last week (November 25 to December 1) and the penultimate week (November 18 to 24). In the example shown in Figure 1C, self-reported work hours were mostly lower than the app-recorded counterparts across time, by an average of 1.19 hours and 1.01 hours every day of the last week and the penultimate week, respectively.

In addition, the percentage of days that participants were unable to recall work hours precisely (NR) was recorded. If the

participants reported they were unable to recall the work hours of a particular day, the interviewers provided cues by offering them their average time of arriving and leaving work in the last month according to the participant's report earlier during the interview. Self-reported work hours of the previous week (WH_{self-1}) and the penultimate week (WH_{self-2}) were calculated by summing the reported work hours of each day.

Recall Bias Indicators

We used two groups of indicators to quantify recall bias. The first group included the differences between self-reported and app-recorded work hours of the previous week (D_1), penultimate week (D_2), and previous month (D_M), in which D is defined as $WH_{self}–WH_{app}$. The second group included the percentage of days that participants were unable to precisely recall their work hours of the previous week (NR_1) and the penultimate week (NR_2) during interviews.

Statistical Analysis

We compared the self-reported work hours with their app-recorded counterparts with a paired *t* test, and examined their correlation with Pearson correlation coefficients. We compared the percentage of days that participants were unable to recall their work hours precisely for the previous week (NR_1) and the penultimate week (NR_2) with a paired *t* test. We also used Pearson correlation coefficients to examine the associations between work hours and the recall bias indicators D_1, D_2, NR_1, and NR_2. In addition, we examined the test-retest reliability of the recall bias indicators by intraclass correlation coefficients between the first month and the second month; $P<.05$ was considered to indicate statistical significance. Data arrangement and statistical analysis were performed using IBM SPSS Statistics 25.

Results

An average of 9.7 (SD 3.2) minutes was required to complete a participant's interview every month. The percentage of days that participants were unable to precisely recall their work hours of the penultimate week (NR_2; mean 38.6%, SD 33.9%) was significantly higher than that of the previous week (NR_1; mean 16.0%, SD 23.7%) by 18.5% ($P=.004$). The standard deviation of day-to-day work hours was 3.8 hours.

Both WH_{app-1} and WH_{app-2} presented a normal distribution according to the Kolmogorov-Smirnov test ($P=.20$ for both) and Shapiro-Wilk tests ($WH_{app-1}P=.26$, $WH_{app-2}P=.77$). These self-reported work hours were highly correlated to their app-recorded counterparts for the previous week ($r=–0.87$, $P<.001$), penultimate week ($r= 0.88$, $P<.001$), and previous month ($r=0.86$, $P<.001$). Table 1 shows that the self-reported hours were significantly lower than their app-recorded counterparts, with the greatest average differences for the previous week (D_1), followed by the penultimate week (D_2), and the smallest difference found for the previous month (D_M) (all $P<.001$).

Table 1. Comparison of self-reported and app-recorded work hours of the previous week, the penultimate week, and previous month.

Time period recorded	Self-reported work hours	App-recorded work hours	Difference, mean (SD)	*P* value
Previous week	57.30	65.63	–8.33 (8.95)	<.001
Penultimate week	58.57	65.65	–7.08 (8.74)	<.001
Previous month (weekly)	60.26	66.94	–6.68 (8.27)	<.001

Table 2 shows the correlation coefficients between the recall bias indicators (D_1, D_2, NR_1, NR_2) and the app-recorded work hours (WH_{app-1} and WH_{app-2}). WH_{app-1} was significantly negatively correlated to D_1, meaning that the longer WH_{app-1}, the more negative the difference between WH_{self-1} and WH_{app-1}, representing more underestimation of self-reported work hours. WH_{app-1} was not significantly correlated to NR_1. By contrast, WH_{app-2} was significantly correlated to NR_2, meaning that the longer WH_{app-2}, the higher percentage of days that work hours were not precisely recalled during interviews. WH_{app-2} was not correlated to D_2 and WH_{app-m} was not significantly correlated to D_M.

Table 2. Pearson correlation coefficients (r) of app-recorded work hours and recall bias indicators.

Recall bias indicator	WH_{app-1}[a]		WH_{app-2}[b]	
	r	*P* value	*r*	*P* value
Difference between self-reported and app-recorded work hours	–0.41	.01	–0.12	.48
Percentage of days that participants were unable to recall their work hours precisely	0.08	.37	–0.49	.002

[a]app-recorded work hours of the previous week.

[b]app-recorded work hours of the penultimate week.

Discussion

Principal Findings

This study demonstrated the underestimation of 11.5%-12.0% (D_1/WH_{app-1} and D_2/WH_{app-2}) work hours and the association between this recall bias and excessive work hours, with the novel app "Staff Hours" recording GPS-defined work hours. We recruited a total of 18 medical doctors (11 men; median age 29 years, range 24-53 years) as participants. As methodological strengths, this study helps to advance the field of work-hours estimation by recording work hours in efficient, precise ways and with higher temporal resolution. The app automatically recorded work hours in real time, and an average of 9.7 minutes was required to complete a participant's interview every month. We were able to collect self-reported data with structured interviews twice per participant, using psychiatrists as interviewers. An earlier study conducted in the United States in 2004 including 45 female flight attendants as participants examined work hours per month [13]. Another study conducted in Japan in 2016 included 164 male employees as participants to gather work-hour information with questionnaires [14]. Studies using self-reported questionnaires typically processed work-hour details as ordinal variables [14], such as by recording the work hours roughly (eg, 45 to <60 hours per week, 60 to <80 hours per week, and >80 hours per week) rather than as continuous variables (eg, 67 hours per month), although the latter has higher precision than using ordinal variables. In addition, the app recorded data with higher temporal resolution than obtained from self-reports. This study included 18 participants and ran over 61 work days, contributing 1068 pairs of app-recorded and self-reported data after subtracting missing data. On the basis of the high temporal resolution of the data, the standard deviation of the average day-to-day work hours was 3.8 hours, highlighting the fluctuating nature of physicians' work hours.

Our result that app-recorded work hours strongly correlated with their self-reported counterparts (r=0.86-0.88, P<.001) was consistent with previous studies [2,14]. One previous study [14] also demonstrated a high correlation between self-reported and actual work hours. Our study further extends this previous research by describing the extent of underestimation; the weekly underestimation was 6.48 to 8.97 work hours per week, which represents approximately 1.30-1.79 hours of underestimation every day considering a 5-day work week (6.48/5=1.30; 7.05/5=1.79). Differences of temporal resolution between the app-recorded work hours and participants' recall were noted, which may contribute to systematic biased recall. When participants were unable to recall the work hours of a particular day precisely, they tended to report rounded-up hours, typically to the nearest half hour (eg, reporting as having arrived at work at 6:30 AM or 7:00 AM), as well as reporting regular work hours, which were often lower than their actual work hours. The app scanned at 10-minute intervals to determine the GPS coordinates and therefore the initiation and cessation of recording work hours. Our results showed that all participants reported starting at the scheduled work time, despite actually having arrived at the workplace a few minutes earlier than their work hours according to GPS coordinates.

Previous studies [9,10] showed that when duty hour regulations were enforced, resident physicians were faced with a dilemma between violating the duty hour regulations and maintaining patient care quality. When facing such a dilemma, an option that resident physicians chose was working overtime while still reporting the duty hours within the limitations of regulations. It was estimated that up to 60% of surgical residents reported their work hours untruthfully when filling out questionnaires. The reasons resident physicians underreported their work hours might have included pressure from the system (eg, receiving punishment due to violations of regulations) and pressure from peers (eg, being viewed as "incompetent" when leaving work on time).

Besides differences in temporal resolution between app-recorded and self-reported work hours, distorted time perception resulting from long work hours may also contribute to recall bias, as demonstrated by the two recall bias indicators used in this study. Physicians whose work hours were longer during the previous week demonstrated a tendency toward more underestimation during recall (r=–0.410, P=.01), whereas participants whose work hours were longer in the penultimate week showed a diminished tendency to underestimate their work hours during recall (r=–0.119, P=.48). In addition to previous studies that simply showed the extent to which recall of work hours was impaired [2,14], our study recorded the percentage of work days that participants were unable to recall their work hours. Participants whose work hours were longer during the penultimate week showed a lower rate of recalling their work hours precisely when compared with those who worked shorter hours (r=0.489, P=.002). As a result of a higher NR, the interviewers provided cues more frequently when participants recalled their work hours of the penultimate week, thus allowing for attenuation of underestimation.

Excessive work hours may distort the time perception and memory of the extent of work hours. Our previous studies performed in 2012 involving 74 medical interns showed that physicians in Taiwan worked an average of 86.7 hours per week with up to 33.5 consecutive working hours per duty shift, and that they sometimes developed hypervigilant perceptions, phantom vibration, and ringing-ear syndrome in the absence of an external stimulus [1,15]. Excessive work hours also resulted in reduced cardiac sympathetic modulation [1,16], disrupted sleep stability [17], and increased anxiety and depression symptoms [18-20]. These psychological and physiological impacts of excessive work hours could alter time memory and time perception (ie, the ability to recall work hours), thereby increasing recall bias [21,22].

Limitations

This study has several methodological limitations that should be considered when interpreting the results. First, if the participants reported that they were unable to recall the work hours of a particular day, the interviewers provided cues by offering them their average time of arriving and leaving work in the last month. For some participants who were more unable to recall their work hours (ie, those with a larger NR), their D values were calculated based more on cues provided by the interviewers. By contrast, NR was not affected by any cues

provided by the interviewers. Therefore, NR might be a better self-defined recall bias indicator compared with D. Second, the participants were not asked to recall any work hours earlier than the penultimate week, which thus limited the studied timeframe of recall bias in this study. However, the percentage of days that participants were unable to recall work hours of the penultimate week was 38.6%. It was conceivable that reports of work hours earlier than 2 weeks prior may result in greater recall bias. Third, the sample size of this study was relatively small, which limits the generalization of the findings as well as the detection of other recall bias indicators. However, we were able to ensure participant diversity, and the 2-month length of app installation guaranteed adequate data input for analysis. Future research should investigate how a larger sample size would identify more factors associated with recall bias indicators. Future studies are also warranted to assess the underlying mechanisms of time perception and recall bias of work hours.

Conclusion

This pilot study identified the existence of recall bias of work hours, the extent to which the recall was biased (6.48-8.97 work hours per week), and the influence of work hours on recall bias. We were able to demonstrate that long work hours in the previous and the penultimate week influenced recall bias in different ways. Working overtime has long been part of the culture among Taiwanese physicians [1,15,17,19,20,23]. However, the extent to which medical staff overwork has never been systemically investigated, and hence the work-hour regulations of physicians were formed without the basis of solid work-hour studies. This study has clear public health relevance by confirming that this time-saving, easily accessible app can provide solid, precise reports of actual work hours, aid work policy formation for physicians, and promote the care quality of patients, as well as the well-being of physicians in Taiwan.

Acknowledgments

The authors thank the Taipei Doctors Union and Ting-Chen Hsu for their excellent technical assistance. This study was supported by grants from the Ministry of Science and Technology of Taiwan (MOST 110-2314-B-400-054).

Authors' Contributions

HW and YL designed the study, performed the study, analyzed the data, and drafted and revised the manuscript. Both authors agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Lin Y, Kuo TBJ, Ho Y, Lin S, Liu C, Yang CCH. Physiological and psychological impacts on male medical interns during on-call duty. Stress 2012 Jan 19;15(1):21-30. [doi: 10.3109/10253890.2011.572208] [Medline: 21682655]
2. Lockley SW, Cronin JW, Evans EE, Cade BE, Lee CJ, Landrigan CP, Harvard Work Hours, HealthSafety Group. Effect of reducing interns' weekly work hours on sleep and attentional failures. N Engl J Med 2004 Oct 28;351(18):1829-1837. [doi: 10.1056/NEJMoa041404] [Medline: 15509816]
3. Landrigan CP, Fahrenkopf AM, Lewin D, Sharek PJ, Barger LK, Eisner M, et al. Effects of the accreditation council for graduate medical education duty hour limits on sleep, work hours, and safety. Pediatrics 2008 Aug 01;122(2):250-258. [doi: 10.1542/peds.2007-2306] [Medline: 18676540]
4. Bilimoria KY, Chung JW, Hedges LV, Dahlke AR, Love R, Cohen ME, et al. National cluster-randomized trial of duty-hour flexibility in surgical training. N Engl J Med 2016 Feb 25;374(8):713-727. [doi: 10.1056/NEJMoa1515724] [Medline: 26836220]
5. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. J Grad Med Educ 2015 Sep;7(3):349-363 [FREE Full text] [doi: 10.4300/JGME-D-14-00612.1] [Medline: 26457139]
6. Drolet BC, Sangisetty S, Tracy TF, Cioffi WG. Surgical residents' perceptions of 2011 Accreditation Council for Graduate Medical Education duty hour regulations. JAMA Surg 2013 May 01;148(5):427-433. [doi: 10.1001/jamasurg.2013.169] [Medline: 23677406]
7. Landrigan CP, Rahman SA, Sullivan JP, Vittinghoff E, Barger LK, Sanderson AL, et al. Effect on patient safety of a resident physician schedule without 24-hour shifts. N Engl J Med 2020 Jun 25;382(26):2514-2523. [doi: 10.1056/nejmoa1900669]
8. Ogawa R, Seo E, Maeno T, Ito M, Sanuki M, Maeno T. The relationship between long working hours and depression among first-year residents in Japan. BMC Med Educ 2018 Mar 27;18(1):50 [FREE Full text] [doi: 10.1186/s12909-018-1171-9] [Medline: 29587738]
9. Sticca RP, Macgregor JM, Szlabick RE. Is the Accreditation Council for Graduate Medical Education (ACGME) resident/fellow survey, a valid tool to assess general surgery residency programs compliance with work hours regulations? J Surg Educ 2010 Nov;67(6):406-411. [doi: 10.1016/j.jsurg.2010.09.007] [Medline: 21156299]

10. Fargen KM, Rosen CL. Are duty hour regulations promoting a culture of dishonesty among resident physicians? J Grad Med Educ 2013 Dec;5(4):553-555 [FREE Full text] [doi: 10.4300/JGME-D-13-00220.1] [Medline: 24454999]
11. Chiang T, Chen S, Pan Y, Lin Y. Automatic work-hours recorder for medical staff (Staff Hours): mobile app development. JMIR Mhealth Uhealth 2020 Feb 25;8(2):e16063 [FREE Full text] [doi: 10.2196/16063] [Medline: 32130165]
12. Medical residents' labor rights protection guideline. Ministry of Health and Welfare. 2017. URL: https://dep.mohw.gov.tw/DOMA/cp-5050-57749-106.html [accessed 2019-10-25]
13. Grajewski B, Atkins DJ, Whelan EA. Self-reported flight hours vs. company records for epidemiologic studies of flight attendants. Aviat Space Environ Med 2004 Sep;75(9):806-810. [Medline: 15460634]
14. Imai T, Kuwahara K, Miyamoto T, Okazaki H, Nishihara A, Kabe I, et al. Validity and reproducibility of self‐reported working hours among Japanese male employees. Jrnl of Occup Health 2016 Jul 22;58(4):340-346. [doi: 10.1539/joh.15-0260-oa]
15. Lin Y, Chen C, Li P, Lin S. A dimensional approach to the phantom vibration and ringing syndrome during medical internship. J Psychiatr Res 2013 Sep;47(9):1254-1258. [doi: 10.1016/j.jpsychires.2013.05.023] [Medline: 23786911]
16. Chien J, Chen C, Lin S, Lin S, Lin Y. Cardiac autonomic modulation during on-call duty under working hours restriction. Int J Environ Res Public Health 2020 Feb 10;17(3):1118 [FREE Full text] [doi: 10.3390/ijerph17031118] [Medline: 32050580]
17. Lin Y, Ho Y, Lin S, Yeh Y, Liu C, Kuo TBJ, et al. On-call duty effects on sleep-state physiological stability in male medical interns. PLoS One 2013 Jun 4;8(6):e65072 [FREE Full text] [doi: 10.1371/journal.pone.0065072] [Medline: 23750232]
18. Chen C, Lin S, Li P, Huang W, Lin Y. The role of the harm avoidance personality in depression and anxiety during the medical internship. Medicine (Baltimore) 2015 Jan;94(2):e389. [doi: 10.1097/MD.0000000000000389] [Medline: 25590843]
19. Lin Y, Chen C, Lin S, Liu C, Weng W, Kuo TBJ, et al. Gender differences in cardiac autonomic modulation during medical internship. Psychophysiology 2013 Jun 20;50(6):521-527. [doi: 10.1111/psyp.12038] [Medline: 23517383]
20. Lin Y, Chen H, Tsai S, Chang L, Chen P. A prospective study of the factors associated with life quality during medical internship. PLoS One 2019 Aug 12;14(8):e0220608 [FREE Full text] [doi: 10.1371/journal.pone.0220608] [Medline: 31404080]
21. Lake JI, Labar KS, Meck WH. Corrigendum to "Emotional modulation of interval timing and time perception" [Neurosci. Biobehav. Rev. 64 (2016) 403-420]. Neurosci Biobehav Rev 2020 Sep;116:182. [doi: 10.1016/j.neubiorev.2020.04.027] [Medline: 32623151]
22. Virtanen M, Singh-Manoux A, Ferrie JE, Gimeno D, Marmot MG, Elovainio M, et al. Long working hours and cognitive function: the Whitehall II Study. Am J Epidemiol 2009 Mar 01;169(5):596-605 [FREE Full text] [doi: 10.1093/aje/kwn382] [Medline: 19126590]
23. Lin Y, Lin S, Li P, Huang W, Chen C. Prevalent hallucinations during medical internships: phantom vibration and ringing syndromes. PLoS One 2013 Jun 10;8(6):e65152 [FREE Full text] [doi: 10.1371/journal.pone.0065152] [Medline: 23762302]

Abbreviations

ACGME: Accreditation Council for Graduate Medical Education
D: difference between self-reported and app-recorded work hours
NR: percentage of days that work hours were not precisely recalled during interviews
WH: work hours

Edited by R Kukafka; submitted 24.12.20; peer-reviewed by LA Lee, H Mehdizadeh, JR Bautista; comments to author 05.04.21; revised version received 10.05.21; accepted 08.11.21; published 24.12.21.

Please cite as:
Wang HH, Lin YH
Assessing Physicians' Recall Bias of Work Hours With a Mobile App: Interview and App-Recorded Data Comparison
J Med Internet Res 2021;23(12):e26763
URL: https://www.jmir.org/2021/12/e26763
doi:10.2196/26763
PMID:34951600

Original Paper

A Comparison of the Use of Smart Devices, Apps, and Social Media Between Adults With and Without Hearing Impairment: Cross-sectional Web-Based Study

Marieke F van Wier[1], PhD; Emily Urry[2], PhD; Birgit I Lissenberg-Witte[3], PhD; Sophia E Kramer[1], PhD

[1]Ear & Hearing, Otolaryngology-Head and Neck Surgery, Amsterdam Public Health research institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Research & Development, Sonova AG, Stäfa, Switzerland

[3]Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Marieke F van Wier, PhD
Ear & Hearing, Otolaryngology-Head and Neck Surgery
Amsterdam Public Health research institute
Amsterdam UMC, Vrije Universiteit Amsterdam
De Boelelaan 1117
Amsterdam, 1081 HV
Netherlands
Phone: 31 20 4440952
Email: m.vanwier1@amsterdamumc.nl

Abstract

Background: eHealth and social media could be of particular benefit to adults with hearing impairment, but it is unknown whether their use of smart devices, apps, and social media is similar to that of the general population.

Objective: Our aim is to study whether adults with normal hearing and those with impaired hearing differ in their weekly use of smart devices, apps, and social media; reasons for using social media; and benefits from using social media.

Methods: We used data from a Dutch cohort, the National Longitudinal Study on Hearing. Data were collected from September 2016 to April 2020 using a web-based questionnaire and speech-in-noise test. The results from this test were used to categorize normal hearing and hearing impairment. Outcomes were compared using (multiple) logistic regression models.

Results: Adults with impaired hearing (n=384) did not differ from normal hearing adults (n=341) in their use of a smartphone or tablet. They were less likely to make use of social media apps on a smartphone, tablet, or smartwatch (age-adjusted odds ratio [OR] 0.67, 95% CI 0.48-0.92; *P*=.02). Use of social media on all devices and use of other apps did not differ. Adults with hearing impairment were more likely to agree with using social media to stay in touch with family members (OR 1.54, 95% CI 1.16-2.07; *P*=.003) and friends (age-adjusted OR 1.35, 95% CI 1.01-1.81; *P*=.046). Furthermore, they were more likely to agree with using social media to perform their work (age-adjusted OR 1.51, 95% CI 1.04-2.18; *P*=.03). There were no differences in the experienced benefits from social media.

Conclusions: The potential for eHealth is confirmed because adults with hearing impairment are not less likely to use smart devices than their normal hearing peers. Adults with hearing impairment are less likely to use social media apps on a smart device but not less likely to use social media on all types of internet-connected devices. This warrants further research on the types of social media platforms that adults with hearing impairment use and on the type of device on which they prefer to use social media. Given that participants with hearing impairment are more likely than their normal hearing peers to use social media to perform their work, use of social media may be seen as an opportunity to enhance vocational rehabilitation services for persons with hearing impairment.

(J Med Internet Res 2021;23(12):e27599) doi:10.2196/27599

KEYWORDS

hearing impairment; social media use; app use; benefits from social media; eHealth; mobile phone

Introduction

Background

Disabling hearing loss affects 466 million people worldwide [1]. Its prevalence will double by 2050 because of increasing life expectancy. Hearing impairment is one of the most prevalent disabilities because hearing deteriorates with age. Almost one-third of people aged ≥65 years have disabling hearing loss [1]. However, because age-related hearing loss can start earlier, a large number of middle-aged adults also have hearing disabilities. In high-income countries, 15%-25% of adults between the ages of 50 and 65 years have mild to severe hearing loss [2,3]. A recent study in the Netherlands estimated that 1.2 million adults aged ≥40 years, or 13% of the population in that age bracket, have disabling hearing loss [4]. Hearing difficulties can be mitigated by the use of communication strategies, speech reading, hearing aids, and hearing assistive technology. A smaller number of people with impaired hearing, mostly those who have been deaf or hard of hearing from a young age, communicate in sign language as their native language.

Most adults in high-income countries have access to the internet, although its use is less ubiquitous in older adults [5,6]. This gives potential for digitalized hearing health care, both stand-alone and adjunct to in-person care [7-9]. Digital hearing health care facilitates patient-centered care in the comfort of one's own home. In this way, next to making hearing health care more accessible, digital care could perhaps boost the lagging uptake of communication strategies and hearing aids [10]. The current COVID-19 pandemic has sped up the shift to remote audiological care, although audiologists still have concerns about patients' access to technology and their preferences [11]. During the UK lockdown, teleaudiology was particularly used for (tinnitus) counseling [11]. The internet offers many synchronous and asynchronous communication options that support counseling, for example, email, direct messaging, social network sites and apps, and video calling. With the development of web-based hearing assessment and hearing aid fitting, these in-person services will also be offered on the web [12]. In addition to teleaudiology, internet-mediated communication could also support psychosocial health by fostering social connection. Mild to severe hearing loss makes it difficult to follow conversations in certain situations, even with the use of a hearing aid, leading to less meaningful interactions and withdrawal from social activities [13]. Loneliness, depression, and anxiety are more prevalent in adults with hearing loss than in their peers who have no hearing problems [13-16]. Communication and connection with others through the internet could replace some of the face-to-face contacts and mitigate these negative outcomes [17]. This raises the question of whether adults with hearing impairment use the internet, smart devices, apps, and social media as much and for the same reasons as their normal hearing peers and the general public. If so, it would be reasonable to move (some) services to the web, which would substantiate the opportunities for social support. Earlier research disputes the assumption of equal use.

So far, only 3 studies have looked at internet access and use among adults with hearing impairment in high-income countries. Of these studies, 2 [18,19] showed that older adults with hearing difficulty were less likely to use the internet, whereas the other [20] found no differences between users of hearing aids and the general population overall. However, the latter study did find a higher internet use among users of hearing aids aged ≥75 years compared with the (age-matched) general population. Thus, there is inconclusive evidence regarding whether adults with hearing loss make different use of the internet than their normal hearing peers. Moreover, these studies were performed at a time when the internet was mostly accessed through desktop computers and was less pervasive in daily life. A more recent marketing survey looked at whether ownership and use of a smartphone is different in adults with a hearing impairment [21]. This 2018 UK survey found that the rates of smartphone ownership were 53% in people with self-reported hearing impairment and 81% in people who reported having no disabilities. Furthermore, 44% of the adults with hearing impairment used their mobile device to access the internet, whereas this rate was 69% in respondents without disabilities. However, 60% of the people who reported a hearing impairment were aged >65 years compared with only 16% of the nondisabled people. The results might therefore reflect the differences in age between both groups. The same survey found that people with a hearing impairment had quite similar reasons for using the internet as the people who reported no disability. Comparable results were obtained from a 2016 US survey [22]. This survey asked people with self-reported hearing disabilities which features or functions they use on their smartphone. Use of the internet in general and of social media apps in particular seems to be a little lower in people with a hearing disability compared with the general population, but as in the UK survey, this may also reflect differences in age.

Several studies have analyzed social media use by distinct groups of adults with hearing impairment by extracting publicly available content and user data (manual and automated *web scraping*) posted on social network sites such as Facebook, Twitter, and YouTube, as well as from web-based forums and from personal blogs [23-28]. This type of research gives insight into the information that is shared, the feelings and opinions of the posters, and the nature of communication on these platforms, but there are also limitations. Content analysis studies do not give insight into either the relative frequency of social media use or its passive use (ie, reading and watching). Furthermore, the characteristics of the posters, such as age, sex, income, education level, and whether they have hearing loss and to what extent, are unknown or not verifiable. This impedes specification of who uses which type of social media and in which way and who does not. In addition, there is no (legal) access to private communication such as WhatsApp messages and private Facebook pages. Postings on social media also do not provide information about the potential benefits of using social media, for example, whether it has strengthened the social connections of the user. In conclusion, these studies do not give insight into the social media use of all adults with hearing impairment: whether they use social media for private communication, the extent of use, and their reasons for using social media; the experienced benefits from using social media; and whether all these data are comparable with those found in the general public and normal hearing adults. The latter would justify relying on

data from the general public when making decisions on using computer-mediated communication with adults with hearing impairment. We did not find studies that addressed these questions.

Research Questions

This study therefore investigated the following research questions (RQs):

- RQ1: Do adults with normal hearing and those with impaired hearing differ in the use of smart devices?
- RQ2: Do adults with normal hearing and those with impaired hearing differ in the use of different types of apps?
- RQ3: Do adults with normal hearing and those with impaired hearing differ in the use of social media?
- RQ4: Do adults with normal hearing and those with impaired hearing differ in their reasons for using social media?
- RQ5: Do adults with normal hearing and those with impaired hearing differ in the experienced benefits from using social media?

Methods

Overview

Data were available from a long-running Dutch national cohort of adults with normal hearing and those with impaired hearing, the National Longitudinal Study on Hearing (NL-SH). Demographics and use of technology, apps, and social media were collected through a web-based questionnaire. The hearing status of participants was determined by a web-based speech-in-noise test.

Details of NL-SH

Recruitment and Measurements

Initiated in 2006, the NL-SH is an ongoing, prospective cohort study in which both adults with normal hearing and those with impaired hearing participate. This cohort was set up to gain knowledge on the long-term trajectory of hearing loss and its association with psychosocial health, work outcomes, and health care use in adults of working age. The NL-SH uses convenience sampling. The major portals for recruitment are the web-based Dutch National Hearing Test (NHT) [29] and the study website [30] where the public can take the same hearing test. It offers the general public a fast and convenient way to test their own ability to recognize speech in noise (further described in the *Speech-in-Noise Test* section). After presenting the test results, they are asked if they are interested in taking part in hearing-related research. If they are interested, they are taken to the introduction page of the NL-SH on the study website [30], where the study is fully explained and they can download an information brochure. Prospective participants can then choose to enroll. At enrollment, sex, age, and contact details (email address, home address, and phone number) are asked. Age is checked against the inclusion criterion of being 18-70 years; those who do not fit this criterion cannot enroll. Other eligibility criteria are not set for the NL-SH. Those who are eligible have to take a speech-in-noise test specific to the NL-SH and are sent a link to the web-based study questionnaire.

Inclusion measurements (T0) started in 2006 and still continue. The 5-year follow-up (T1) started in 2011, and the 10-year follow-up (T2) started in 2016. For this study, only the T2 measurements were used. The invitation for the T2 measurement round was sent approximately 10 years after the T0 hearing test was performed. In all, 2 email reminders and a postal reminder were sent within a period of 3 months after the first invitation to fill out the questionnaire. In general, the T2 hearing test is performed directly after participants fill out the web-based questionnaire, but a delay of up to 3 months is possible. People who did not perform this test received a total of 3 email reminders during these 3 months. To be included in this study, both the T2 questionnaire and hearing test had to be fully completed. Data collected up to April 1, 2020, were included in the analyses.

The NL-SH study protocol has been approved by the Medical Ethics Committee of the Amsterdam Medical Center, location VUmc, in Amsterdam, the Netherlands (METC number: 2006/83; ToetsingOnline NL12015.029.06).

Questionnaire

The questionnaire consisted of questions on demographics; hearing; use of technology, apps, and social media; and reasons for using social media and benefits of this use. The questions on use of technology and apps were taken from a web-based marketing study (Sonova AG, written communication, 2019). The validity of these questions is unknown. Standardized questionnaires about use of social media, reasons for using social media, and the experienced benefits from using social media are not available. The questions on these topics were devised by the research team and pilot-tested in a group of adults with hearing impairment. Textbox 1 lists the questions other than demographics. The full questionnaire can be found in Multimedia Appendix 1.

Textbox 1. Overview of the questions on use of technology, apps, and social media, as well as reasons for using social media and benefits of social media use.

Questions and answer options used in this study

- Which of these devices do you use at least once a week?
  - Smartphone
  - Smartwatch
  - Tablet
- Which types of apps do you use at least once a week on your current smartphone, tablet, or smartwatch?
  - Weather
  - News
  - Finances (mobile banking, stock exchange, etc)
  - Navigation
  - Remote control (television, stereo system, etc)
  - Fitness
  - Communication (email, WhatsApp, WeChat, etc)
  - Medical or health
  - Social media (Facebook, Instagram, Twitter, etc)
  - Music and podcasts
- Do you use social media?
  - Yes
  - No
- To what extent do you agree with the following statements (rated on a scale of 1-10: fully disagree=1, fully agree=10)? I use social media to...
  - Stay in touch with family members
  - Stay in touch with acquaintances
  - Stay in touch with colleagues or peers
  - Share experiences, videos, or photos
  - View experiences, videos, or photos
  - Expand my work-related network
  - Perform my work
  - Gain new knowledge
  - File complaints and problems with the government or businesses
- With what frequency do you use social media?
  - Multiple times a day
  - Daily
  - Weekly
  - Monthly
  - A couple of times a year
- What have you gained from your social media use so far?
  - New acquaintances
  - New friendships
  - Closer or more intense family ties
  - Closer or more intense friendships

- Expanded work-related network
- New knowledge about health
- I have gained little or nothing from social media

Demographics concerned sex, current age, highest attained education, and first language Dutch or other. Education was divided into low (elementary school or attended high school but no degree), medium (high school graduate or having an associate's degree), and high (having a bachelor's degree, master's degree, or doctoral degree). In all, 2 hearing-related questions asked about having normal hearing or some type of hearing loss and, for those with self-reported hearing loss, whether they use a hearing aid.

Use of technology consisted of a list of devices with the question "Which of these devices do you use at least once a week?" The list included devices such as a traditional mobile phone, smartphone, traditional wristwatch, smartwatch (eg, Apple watch), fitness watch (eg, Fitbit), laptop or notebook computer, tablet, television, radio, and more. The list did not mention a PC. Participants could tick the box for all devices that applied. For this study, the use of only mobile devices on which an app can be installed (ie, a smartphone, a tablet, or a smartwatch), further referred to as a smart device, was analyzed.

Next, participants who indicated that they used a smart device were asked what types of apps they used at least once a week. The options to tick were as follows: weather, news, finances (mobile banking, stock exchange, etc), navigation, remote control (television, stereo system, etc), fitness, communication (email, WhatsApp, WeChat, etc), medical or health, social media (Facebook, Instagram, Twitter, etc), music or podcasts, and other. The category *other* was not included in the analyses.

Participants were then asked if they use social media, for example, Facebook, LinkedIn, Skype, WhatsApp, Twitter, and Instagram. If they said "yes" or if they had a missing answer on this question, they were asked about frequency of use (1 answer possible): multiple times a day, daily, weekly, monthly, or a couple of times a year. In the analyses, monthly and a couple of times a year were collapsed into 1 category. Those who used social media were asked how much they agreed with several statements about reasons for using social media. The reasons provided were as follows: to stay in touch with family; to stay in touch with friends; to stay in touch with acquaintances; to stay in touch with colleagues or peers; to share experiences, videos, or photos; to view experiences, videos, or photos; to expand work-related network; to perform work; to gain new knowledge; and to file complaints. These questions could be answered on a scale from 0 to 10 (an 11-point Likert scale), with 0=*fully disagree* (coded as 1), 5=*do not agree, do not disagree* (coded as 6), and 10=*fully agree* (coded as 11), resulting in scores of 1-11, to be handled as a continuous outcome.

Finally, social media users were asked what they had gained from social media use so far. They could choose any of the following answers: new acquaintances, new friendships, closer or more intense family ties, closer or more intense friendships, expanded work-related network, new knowledge about health, gained little or nothing, and other. The category *other* was not used in the analyses.

At the end of the questionnaire, a personal link was provided to the web-based speech-in-noise test.

Speech-in-Noise Test

The ability to recognize speech in noise is salient for measuring the disabling effects of hearing impairment because one of the first and major complaints of adults with hearing impairment is difficulty in understanding what is said when there is background noise [31]. To measure speech-in-noise recognition, the procedures of the NHT were followed. First developed for use by phone, the internet version of the NHT was launched in 2005. To ensure comparability to the earlier measurements, the procedures of this original version are still used in the NL-SH measurements. Participants are instructed to perform the test in a quiet room. Users of hearing aids are instructed to perform the test without their hearing aids. All participants are asked to use headphones for the test, but speakers are also allowed. Participants have to indicate which transducer they used. The test is binaural (ie, diotic), and the results are mainly representative of the better ear. A total of 23 digit triplets (eg, 6-2-5) are presented against a background of masking noise in an adaptive manner: the noise level is fixed in the test, and the speech level varies. After each incorrect response, the subsequent triplet is presented at a level higher by 2 dB, increasing the signal-to-noise ratio (SNR) a level higher of 2 dB. If the participant provides a correct response, the subsequent triplet is presented at a level lower by 2 dB. The speech-reception threshold in noise (SRTn) is calculated by taking the average SNR of the last 20 presentations, corresponding to a score of 50% of the presented triplets understood correctly. Because of the design of the test, the SRTn values range from –13.4 dB SNR to a ceiling level of 4 dB SNR.

Validation of the original NHT version showed a high correlation (ρ=0.866) with SRTns derived from the standard test in the Netherlands that uses sentences in stationary speech-shaped noise. The measurement error (SE of measurement) is estimated to be <1 dB [32]. Compared with the standard sentence-in-noise test, the NHT phone version has a sensitivity of 0.91 and a specificity of 0.93 at a cutoff of –4.1 dB for hearing impairment [32]. Because of the benefit of listening with 2 ears, the cutoff of the internet version was adjusted with 1.4 dB [33]. This gives a cutoff of –5.5 dB to divide the group into adults with normal hearing ability to recognize speech in noise (further indicated as normal hearing) and those with impaired hearing. It should be taken into account that diotic speech understanding in noise is less compromised in conductive and mixed hearing losses. This means that participants with these types of hearing loss, even those normally using a hearing aid, may be classified as normal hearing.

Participants using speakers may perform slightly worse in the test than if they had used headphones, resulting in misclassification to the group with hearing impairment for some of the participants with normal hearing.

Statistical Analysis

Normally distributed continuous data are described with means and SDs, whereas nonnormally distributed continuous data are described with medians and IQRs. For nominal and ordinal data, frequencies are reported.

Adults with impaired hearing and normal hearing adults were compared using multiple logistic regression analysis. Age, sex, and education level were considered potential confounders. They were included in the model if (1) the potential confounder had influence (P<.10) on both the outcome and the independent variable and (2) the regression coefficient of the influencing factor changed by ≥10% after adding the potential confounder to the model.

Answers to the items that were scored on an 11-point Likert scale (fully disagree-fully agree) were not normally distributed and had to be categorized. To prevent uneven numbers in the categories, they were categorized into 3 groups based on approximate 1/3 divisions with increasing levels of agreement. Multiple ordinal logistic regression analysis was used to analyze whether the distribution of these answers and answers about the frequency of social media use (several times a day, daily, weekly, monthly, or a couple of times a year) differed between adults with normal hearing and those with hearing impairment. Assumptions of logistic analysis were tested in all analyses. For the ordinal results where the assumption of proportional odds was not met, multinomial logistic regression analysis was used. The results were considered statistically significant if P<.05. Analyses were conducted using SPSS software (base edition with custom tables and advanced statistics add-on; version 26.0; IBM Corporation).

Results

Overview

A total of 885 study participants responded to the T2 measurement round between September 15, 2016, and April 1, 2020. Of the 885 participants, 837 (94.6%) fully filled out the questionnaire. Of these 837 participants, 725 (86.6%) also performed the speech-in-noise test. Of the 112 participants who did not perform the speech-in-noise test, 68 (60.7%) reported in the questionnaire that they had hearing loss.

Of the 725 participants with complete data, 619 (85.3%) participated from September 2016 to December 2017. Most of the participants were women (442/725, 61%). The participants' mean age was 57.7 (SD 11.4) years, and 60% (435/725) had a high level of education (Table 1).

Of the total group, 53% (384/725) had a hearing test score ≥–5.5 dB SNR and were subsequently classified as having a hearing impairment. Table 1 shows the characteristics of these participants and of the normal hearing participants separately.

Table 1. Overview of characteristics and categorical outcomes for the total group, a group with normal hearing, and a group with hearing impairment (N=725).

Characteristics	Total group (N=725)	Normal hearing (n=341)	Hearing impairment (n=384)
Female, n (%)	442 (61)	200 (58.7)	242 (63)
Age (years), mean (SD)	57.7 (11.4)	55.5 (10.9)	59.6 (11.4)
Education level, n (%)			
Low	92 (12.7)	39 (11.4)	53 (13.8)
Medium	198 (27.3)	96 (28.2)	102 (26.6)
High	435 (60)	206 (60.4)	229 (59.6)
First language (Dutch), n (%)	703[a] (98.6)	331 (99.1)	372 (98.1)
Self-reported hearing impairment, n (%)	428 (59)	109 (31.9)	319 (83.1)
Hearing aid use, n (%)	299 (41.2)	46 (13.5)	253 (65.9)
Test with headphones, n (%)	319 (44)	211 (61.9)	108 (28.1)
Weekly use of smart devices, n (%)			
Smartphone	612 (84.4)	298 (87.4)	314 (81.8)
Tablet	420 (57.9)	196 (57.5)	224 (58.3)
Smartwatch	18 (2.5)	8 (2.3)	10 (2.6)
Weekly use of apps,[b] n (%)	660 (100)	313 (100)	374 (100)
Weather	456 (69.1)	217 (69.3)	239 (68.9)
News	453 (68.6)	210 (67.1)	243 (70)
Financial	222 (33.6)	103 (32.9)	119 (34.3)
Navigation	283 (42.9)	132 (42.2)	151 (43.5)
Remote control	148 (22.4)	71 (22.7)	77 (22.2)
Fitness	59 (8.9)	33 (10.5)	26 (7.5)
Communication	551 (83.5)	362 (83.7)	289 (83.3)
Medical or health	82 (12.4)	45 (14.4)	37 (10.7)
Social media	374 (56.7)	198 (63.3)	176 (50.7)
Music and podcasts	175 (26.5)	97 (30.9)	78 (22.5)
I use social media, n (%)	625[c] (86.6)	302 (88.6)	323[d] (84.8)
Frequency of social media use, n (%)	627 (100)	302 (100)	325 (100)
Multiple times a day	310 (49.4)	149 (49.3)	161 (49.5)
Daily	241 (38.4)	118 (39.1)	123 (37.8)
Weekly	56 (8.9)	27 (8.9)	29 (8.9)
Monthly or a couple of times a year	20 (3.2)	8 (2.7)	12 (3.7)
Experienced benefits from social media use, n (%)	628 (100)	302 (100)	326 (100)
New acquaintances	114 (18.2)	59 (19.5)	55 (16.9)
New friendships	72 (11.5)	37 (12.3)	35 (10.7)
Closer or more intense family ties	202 (32.2)	89 (29.5)	113 (34.7)
Closer or more intense friendships	161 (25.6)	78 (25.8)	83 (25.5)
Expanded work-related network	144 (22.9)	81 (26.8)	63 (19.3)
New knowledge about health	141 (22.5)	58 (19.2)	83 (25.5)
Little or no benefit from social media	184 (29.3)	97 (32.1)	87 (26.7)

[a]n=713.

[b]Use of an app on a smartphone, tablet, or smartwatch.

[c]n=722.
[d]n=384.

RQ1: Weekly Use of Smart Devices

Most of the participants made weekly use of a smartphone: 87.4% (298/341) of the normal hearing group and 81.8% (314/384) of the group with impaired hearing (Table 1). Logistic regression analysis, which had to be adjusted for age, revealed that there was no statistically significant difference between the 2 groups in the weekly use of a smartphone (odds ratio [OR] 0.82, 95% CI 0.53-1.25; *P*=.35; Table 2).

Table 2. Odds ratios (ORs) for smart device use, social media use, and experienced benefits from social media use for adults with hearing impairment compared with normal hearing adults; results from (multiple) logistic regression analysis.

Outcome	Crude model		Age-adjusted model[a]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Weekly use of smart devices				
Smartphone	0.65 (0.43-0.977)	.04	0.82 (0.53-1.25)	.35
Tablet	1.04 (0.77-1.39)	.82	N/A[b]	N/A
Weekly use of apps[c]				
Weather	0.98 (0.70-1.36)	.90	0.90 (0.63-1.25)	.50
News	1.15 (0.83-1.59)	.42	N/A	N/A
Financial	1.06 (0.77-1.47)	.71	N/A	N/A
Navigation	1.06 (0.78-1.44)	.73	N/A	N/A
Remote control	0.97 (0.67-1.40)	.88	0.91 (0.63-1.33)	.62
Fitness	0.69 (0.40-1.18)	.17	0.79 (0.45-1.36)	.39
Communication	0.97 (0.64-1.46)	.88	0.99 (0.65-1.52)	.98
Medical or health	0.71 (0.45-1.13)	.15	N/A	N/A
Social media	0.60 (0.44-0.82)	.001	0.67 (0.48-0.92)	.02
Music and podcasts	0.65 (0.46-0.91)	.01	0.73 (0.51-1.05)	.09
I use social media	0.72 (0.47-1.11)	.14	0.90 (0.57-1.42)	.65
Experienced benefits from social media use				
New acquaintances	0.84 (0.56-1.26)	.39	N/A	N/A
New friendships	0.86 (0.53-1.41)	.55	0.93 (0.56-1.54)	.78
Closer or more intense family ties	1.27 (0.91-1.78)	.16	1.21 (0.86-1.71)	.28
Closer or more intense friendships	0.98 (0.69-1.40)	.92	1.10 (0.76-1.59)	.62
Expanded work-related network	0.65 (0.45-0.95)	.03	0.78 (0.53-1.14)	.20
New knowledge about health	1.44 (0.98-2.10)	.06	N/A	N/A
Little or no benefit from social media	0.77 (0.55-1.09)	.14	0.72 (0.51-1.03)	.07

[a]Age in quartiles: Q1: 29-49 years; Q2: 50-59 years; Q3: 60-66 years; and Q4: 67-81 years.
[b]N/A: not applicable; no age adjustment necessary.
[c]Use of an app on a smartphone, tablet, or smartwatch.

The rates of weekly use of a tablet were 57.5% (196/341) in the normal hearing group and 58.3% (224/384) in the group with hearing impairment. Logistic regression analysis showed that these percentages did not differ between the groups (OR 1.3, 95% CI 0.77-1.4; *P*=.82; Table 2).

Because of the small number of people using a smartwatch, we did not test for differences in the use of this device.

RQ2: Weekly Use of Different Types of Apps

Of the 725 participants, 660 (91%) reported using one or more smart devices on which an app could be installed (ie, a smartphone, tablet, or smartwatch). Their weekly use of these apps is shown in Table 1. Use of communication apps such as an email app or WhatsApp was common in the total sample: 83.5% (551/660). Social media apps such as Facebook, Instagram, and Twitter were used weekly by little more than half of the total sample: 56.7% (374/660). Of the normal hearing participants, 63.3% (198/313) made weekly use of social media

apps compared with 50.7% (176/347) of the adults with hearing impairment. After adjustment for age, a statistically significant difference was found between the groups, with adults with hearing impairment 33% less likely to use social media apps (OR 0.67, 95% CI 0.48-0.92; *P*=.02; Table 2). This was the only type of app for which statistically significant differences in weekly use were found.

RQ3: Use of Social Media

Of 722 participants, 625 (86.6%) reported using social media. Use of social media did not differ between adults with normal hearing and those with hearing impairment (age-adjusted OR 0.90, 95% CI 0.57-1.4; *P*=.65; Table 2).

We also looked at the frequency with which social media was used. Most of the participants (551/627, 87.9%) used social media daily or multiple times a day. Of note, of the 627 participants, 2 (0.3%) had a missing score on use of social media, but they answered the question about frequency. The age-adjusted cumulative OR for frequency of use of social media for adults with hearing impairment compared with normal hearing adults was 0.91 (95% CI 0.67-1.2; *P*=.56), showing that there was no statistically significant difference in the frequency of social media use between the groups.

RQ4: Reasons for Using Social Media

Table 3 shows the distribution of the answers for agreements with reasons for using social media. Most of the participants agreed with the statements that they use social media to stay in touch with family members (213/624, 34.1%, score=10-11; 193/624, 30.9%, score=8-9) and with friends (199/625, 31.8%, score=10-11; 227/624, 36.4%, score=8-9). The statement that they use social media to stay in touch with acquaintances was also agreed with by most of the participants, with 31.8% (199/625) scoring 10-11 and 43.4% (271/625) scoring 7-9. Use of social media to expand their work-related network, perform their work, or file complaints and problems was fully disagreed with by most of the participants; 42.1% (257/611), 45.9% (281/611), and 43.5% (263/605), respectively, scored 1 on this question.

Table 3. Reasons to use social media: descriptive outcomes for the answers given on an 11-point Likert scale, divided over 3 percentile groups representing the lowest, middle, and highest levels of agreement for that statement in approximate tertiles.[a]

I use social media to...	Lowest-level agreement		Middle-level agreement		Highest-level agreement	
	Value, n (%)	Value, mean (SD; range)	Value, n (%)	Value, mean (SD; range)	Value, n (%)	Value, mean (SD; range)
Stay in touch with family members	218 (35.9)	4.1 (2.4; 1-7)	193 (31.8)	8.6 (0.5; 8-9)	213 (35.1)	10.8 (0.4; 10-11)
Stay in touch with friends	199 (31.8)	4.6 (2.3; 1-7)	227 (36.3)	8.6 (0.5; 8-9)	199 (31.8)	10.7 (0.5; 10-11)
Stay in touch with acquaintances	186 (30.5)	4.2 (2.1; 1-6)	271 (44.4)	8.3 (0.8; 7-9)	153 (25.1)	10.7 (0.5; 10-11)
Stay in touch with colleagues or peers	202 (33.2)	1.6 (1.0; 1-4)	169 (27.8)	6.2 (0.6; 5-7)	237 (39)	9.2 (1.1; 8-11)
Share experiences, videos, or photos	173 (28)	2.2 (1.3; 1-5)	251 (40.7)	7.0 (0.8; 6-8)	193 (31.1)	9.8 (0.9; 9-11)
View experiences, videos, or photos	207 (33.6)	3.8 (2.1; 1-6)	173 (28.1)	7.6 (0.5; 7-8)	236 (38.3)	9.8 (0.9; 9-11)
Expand my work-related network	257 (42.1)	1.0 (0.0; 1-1)	130 (21.3)	3.1 (1.2; 2-5)	224 (36.7)	7.7 (1.6; 6-11)
Perform my work	281 (46)	1.0 (0.0; 1-1)	112 (18.3)	3.2 (1.1; 2-5)	218 (35.7)	7.8 (1.8; 6-11)
Gain new knowledge	204 (33.7)	1.6 (1.0; 1-4)	234 (38.6)	6.9 (1.0; 5-8)	168 (27.7)	9.7 (0.8; 9-11)
File complaints and problems with the government or businesses	263 (43.5)	1.0 (0.0; 1-1)	142 (23.5)	3.1 (1.1; 2-5)	200 (33.1)	7.5 (1.6; 6-11)

[a]The numbers in each percentile group can be dissimilar because the group cannot be broken up within a specific Likert score. The groups are therefore an approximation of percentile groups based on thirds of the sample.

The analyses of differences between participants with normal hearing and those with impaired hearing showed 3 statistically significant results (Table 4). Ordinal regression analysis revealed that participants with hearing impairment were more likely to score a higher level of agreement, compared with the lowest level, with the statement that they use social media to stay in touch with family members (OR 1.5, 95% CI 1.2-2.1; *P*=.003) and, after correcting for age, to stay in touch with friends (OR 1.4, 95% CI 1.0-1.8; *P*=.046). Multinomial regression analysis, in which ORs are estimated separately for the levels of agreement, showed that participants with hearing impairment were also more likely to have the highest level of agreement (score 6-11 on the agreement scale) with using social media to perform their work (age-adjusted OR 1.5, 95% CI 1.04-2.2; *P*=.03). However, the comparison for the middle level of agreement (score 2-5) did not show a statistically significant relationship (age-adjusted OR 0.98, 95% CI 0.62-1.5; *P*=.93), meaning that normal hearing participants and those with impaired hearing were equally likely to score 2-5 on this statement compared with scoring 1.

Table 4. Cumulative odds ratios (ORs) for being in a higher percentile of agreeing with reasons for using social media for adults with hearing impairment compared with normal hearing adults; results from (multiple) ordinal logistic regression analysis.

I use social media to...	Crude model		Age-adjusted model[a]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Stay in touch with family members	1.55 (1.16-2.07)	.003	N/A[b]	N/A
Stay in touch with friends	1.25 (0.94-1.67)	.13	1.35 (1.01-1.82)	.046
Stay in touch with acquaintances	1.14 (0.85-1.53)	.34	N/A	N/A
Stay in touch with colleagues or peers	1.03 (0.77-1.38)	.84	1.28 (0.94-1.73)	.11
Expand my work-related network	0.85 (0.63-1.14)	.27	1.03 (0.76-1.41)	.83
Perform my work (reference category is lowest-level agreement[c])				
Middle-level agreement	0.83 (0.54-1.29)	.41	0.98 (0.62-1.55)	.93
Highest-level agreement	1.34 (0.94-1.91)	.11	1.51 (1.04-2.18)	.03
Share experiences, videos, or photos	1.14 (0.85-1.53)	.37	N/A	N/A
View experiences, videos, or photos	0.95 (0.71-1.27)	.71	1.05 (0.77-1.42)	.77
Gain new knowledge	1.06 (0.79-1.43)	.68	N/A	N/A
File complaints and problems with the government or businesses	1.17 (0.87-1.58)	.29	1.06 (0.78-1.44)	.71

[a]Age in quartiles: Q1: 29-49 years; Q2: 50-59 years; Q3: 60-66 years; Q4: 67-81 years.

[b]N/A: not applicable; no age adjustment necessary.

[c]Multinomial logistic regression analysis.

RQ5: Experienced Benefits From Using Social Media

Table 1 shows the benefits that participants experienced from using social media. Almost a third (202/628, 32.2%) of the whole group agreed that using social media had given them closer or more intense family ties. Approximately 1 in 4 (161/628, 25.6%) had gained closer or more intense friendships from using social media. New friendships were found by only 11.5% (72/628) of the whole group, and gaining acquaintances was agreed with at a slightly higher rate (114/628, 18.2%). There was also a substantial percentage of participants who had gained little or nothing from social media: 29.3% (184/628).

Logistic regression analysis showed no differences between adults with hearing impairment and those with normal hearing for the experienced benefits from social media use (Table 2).

Discussion

Principal Findings

Adults with hearing impairment and normal hearing adults did not differ in either weekly use of a smartphone or weekly use of a tablet. Adults with hearing impairment were less likely to make weekly use of social media apps on a smartphone, tablet, or smartwatch, but they were not less likely to use social media on all types of devices (including a desktop computer or laptop). Compared with normal hearing adults, adults with hearing impairment were more likely to agree with the statements that they use social media to stay in touch with family members and to stay in touch with their friends. Furthermore, participants with hearing impairment were more likely to be in the group that very much agrees with the statement that they use social media to perform their work. The experienced benefits from social media did not differ between adults with hearing impairment and those with normal hearing.

RQ1: Weekly Use of a Smart Device

Among all NL-SH participants, 84.4% (612/725) made weekly use of a smartphone. This percentage is comparable with the 85% of households in the Netherlands that owned a smartphone in 2017 [34], although it should be noted that our study concerned use of a device and not ownership. After correction for age, we found no differences between adults with normal hearing and those with hearing impairment. Our results therefore are not consistent with those of a 2018 UK survey [21]. This survey found that the rates of smartphone ownership were 53% in people with self-reported hearing impairment and 81% in people who reported having no disabilities. However, 60% of the survey respondents with a self-reported hearing impairment were aged >65 years compared with only 16% of the nondisabled respondents. Although older adults in the United Kingdom are catching up on smartphone use, in 2018 they were still behind younger groups [35]. The results of the UK survey therefore likely reflect the differences in age between both groups.

A tablet is used on a weekly basis by 57.9% (420/725) of all NL-SH participants. In 2017, 66% of Dutch households owned a tablet, which is somewhat higher than in our sample [34]. We found no differences in the use of these devices between adults with normal hearing and those with hearing impairment. No comparisons can be made with other literature because this difference has not been studied before.

RQ2: Weekly Use of Different Types of Apps

Of the 91% (660/725) of the NL-SH participants who use a smart device, 83.5% (551/660) make weekly use of a

communication app such as an email app or WhatsApp. Statistics Netherlands found that 80% of the Dutch general population aged ≥12 years used direct messaging (mostly WhatsApp) in 2017 [34]. This proportion is comparable with the one we found for the weekly use of a communication app. In contrast, Ipsos found a slightly lower percentage for Dutch smartphone users aged 18-64 years [36]. They reported that 68% had used an app to communicate with people in the previous 30 days (data collected in 2017). Use of the following types of apps can be compared with the Ipsos results: weather: NL-SH 69.1% (456/660), Ipsos 54%; news: NL-SH 68.6% (423/660), Ipsos 44%; financial: NL-SH 33.6% (222/660), Ipsos 51%; music and podcasts: NL-SH 26.5% (175/660), Ipsos (listen to music) 37%; and fitness (tracking): NL-SH 8.9% (59/660), Ipsos 20%. Overall, the percentages from the Ipsos report deviate from the ones we found, likely due to differences in the composition of the samples. The mean age of the Ipsos Dutch sample was 40.4 (SD not reported) years (NL-SH 57.7, SD 11.4 years), with 50% of the participants being women (NL-SH: 442/725, 61%) and 78% reporting to be working (NL-SH: 331/725, 45.7%).

We found the rate of overall weekly use of medical and health apps, other than fitness tracking apps, to be 12.4% (82/660). Overall use of these apps among the Dutch population is not known, but in February 2019 the most used nonfitness apps were diet-tracking apps (10.8%), first-aid apps (9.2%), and sleep-tracking apps (8.1%) [37]. These percentages seem to be somewhat higher than those in the NL-SH sample, but this could also be because most of our measurements were performed in 2017. It was only for use of social media apps that we found a difference between adults with normal hearing and those with hearing impairment. After adjustment for age, the adults with hearing impairment had a 33% (OR 0.67, 95% CI 0.48-0.92) lower odds of using social media apps. Another study looked at the use of apps by adults with hearing impairment. A 2016 US survey asked people with self-reported hearing disabilities which features or functions they use on their smartphone and mirrored this with results for all American adults [22]. In the survey, the use of social media apps seems to be lower in people with a self-reported hearing disability (64% vs 75% of all American adults), but no statistical testing was done, nor was an adjustment made for potential age differences. Therefore, it is uncertain if these results agree with those of our study.

It therefore seems that, apart from the use of social media apps, no notable differences exist in the type of apps that adults with hearing impairment use compared with normal hearing adults, but other research should confirm these results.

RQ3: Use of Social Media

Of the NL-SH participants, 86.6% (625/722) used social media on any type of device. Again, this is comparable with the general Dutch population: 85% used social media in 2017 [34]. Most of the NL-SH participants who use social media do this daily or multiple times a day (551/627, 87.9%). This outcome is not available for the general Dutch population, but there is a marketing report describing 72% of WhatsApp users (WhatsApp is the most frequently used social media app in the Netherlands) using this app on a daily basis in 2017 [38]. This is lower than the 87.9% (551/627) we find for social media overall, but this difference is likely caused by this percentage only covering a single social media platform that is also primarily used on a smart device.

Despite the lower weekly use of social media apps by adults with hearing impairment, overall social media use did not differ between adults with normal hearing and those with hearing impairment. A reason for this could be that, in the question on the use of apps, we made a distinction between communication apps (email and direct messaging apps such as WhatsApp) and social media apps (eg, Facebook, Instagram, and Twitter; Textbox 1). The latter addresses use of social networking sites. In the question about overall social media use, direct messaging and social networking sites were taken together. The addition of direct messaging to this question may have obfuscated the difference in use of social networking sites. Another explanation could lie in device preferences. Many social media platforms also run in a web browser. Perhaps adults with hearing impairment have a preference for approaching these social media through a web browser on a laptop or PC because of their larger screens. It is also possible that they prefer to use these devices because they can easily be set up with a speaker. From the hearing test results we know that 71.9% (276/384) of the participants with hearing impairment performed this test with speakers as opposed to using a headphone, whereas this rate was 37.9% (129/340; the type of transducer used was missing for 1 participant) in normal hearing participants. It should also be taken into account that different social media platforms attract users with different characteristics [39]. As we did not ask about distinct social media platforms, we could not compare their specific use. The results could therefore also imply that adults with hearing impairment make less use of social media platforms that are mainly used as an app, for example, Instagram. A final reason for the divergence could be that the use of apps concerned weekly use, whereas use of social media did not specify frequency of use.

We also did not find differences in the frequency of overall social media use. The sparse literature on the latter outcome shows contradictory results and focuses on young people who were born deaf or hard of hearing or became deaf or hard of hearing at a young age [40]. This group is not representative of the participants with hearing impairment in the NL-SH, most of whom have age-related hearing loss. There are important differences in proficiency in reading and writing and in their self-identification between these groups that will likely affect their social media use.

RQ4: Reasons for Using Social Media

A comprehensive theoretical framework on what people do on social media and why they use social media has not been established. Several motivations [41] and theories [42] have been put forward. The activities we asked about in our reasons for using social media mostly fall within the motivations summed up previously [41]. In line with the literature [41,43] and marketing research [44], we found considerable agreement with using social media to keep in touch with people, especially those in the private social network [44]. This is also confirmed by the fact that 80% of the Dutch people aged ≥12 years use

social media for direct messaging. There was also agreement with using social media to connect with colleagues and to create content and consume content, but this was less pronounced. These reasons also concern relationships with other people but are probably less important because the first reason pertains to professional contacts and the latter 2 concern less direct ways of connecting with others. Although 30% of the population of the Netherlands were said to use a professional social network in 2017 [34], daily use of LinkedIn is limited [38]. This confirms the disagreement we found with the statements about social media and work. NL-SH participants also do not agree with filing complaints and problems (with businesses and organizations) using social media. Marketing research in 2017 shows similar outcomes: 9% of Dutch people contacted customer services through WhatsApp, and only 12% of them had a preference for this [45]. There is a major preference for traditional contact through phone and email. Agreement with using social media to gain new knowledge was fairly evenly divided across the participants. The literature also mentions this as a motivation for using social media but gives no clues about its relative importance for using social media [41].

In the comparison between the 2 groups, we found that participants with hearing impairment are more likely to agree with using social media to stay in touch with family members and friends. They were also more likely to have the highest level of agreement with using social media to perform their work compared with the lowest level of agreement. The results of research among young people who are deaf and hard of hearing show that the most frequent motive to use social media is to maintain social contact, although no comparisons were made with normal hearing people [40].

No other studies could be found on use of social media for work purposes among adults with hearing impairment. Thus, we are the first to report that agreement with using social media to keep in contact with family and friends and to perform work is higher among adults with hearing impairment than among normal hearing peers.

RQ5: Experienced Benefits From Using Social Media

Gains or benefits experienced from using social media are conflated with the reasons to use social media. The finding that almost a third (202/628, 32.2%) of the whole group agreed that using social media had given them closer or more intense family ties and that 25.6% (161/628) had gained closer or more intense friendships ties in with many participants using social media to stay in touch with people close to them. Research in adolescents shows that social media can have a positive effect on their social connectedness and sense of belonging [46]. Positive results on social well-being are also found in adults but only when there is low emotional investment [47].

We found no differences between adults with hearing impairment and those with normal hearing for the experienced benefits from social media use. This suggests that despite our previous finding that participants with hearing impairment are more likely to use social media to stay in touch with family members and friends and to perform their work, they derive benefits from this similar to those derived by normal hearing adults.

Implications for Clinical Practice and Research

The broad use of smart devices in our sample shows the potential for digitalized hearing health care through specific apps. However, it should be noted that 9% (65/725) of the NL-SH participants did not use a smart device and that this was more prominent in older adults. Of all study participants aged >65 years, 15.7% (182/216) did not use a smart device (results not shown). As most of the people with hearing impairment are aged >65 years, this means that a substantial number of people with hearing impairment access the internet by other means. To ensure equal access, digital hearing health care should not be restricted to apps; it should also be available on desktop computers and laptops. This ties in with the possibility that adults with hearing impairment have a preference to use these latter devices to access social media.

The use of medical and health apps appears low in our sample because only 12.4% (82/660) reported using these apps and their use is apparently higher in the general public. This does not imply that adults with hearing impairment have little interest in apps for hearing health care; rather, it means that more research is needed into their needs and wants. A variety of solutions may be necessary, as was shown in previous research on eHealth alongside the customer journey of older people with hearing loss [48].

We speculated that communication through the internet could be specifically attractive to adults with hearing impairment. Indeed, adults with hearing impairment were more likely to use social media to stay in touch with family members and friends than normal hearing adults. Nonetheless, they were not more likely to report closer or more intense family ties or friendships from using social media. This shows that internet-mediated communication does not have more salience for adults with hearing impairment than other modes of communication. More research is needed on the type of social media that adults with hearing impairment use to stay in touch with close family members and friends and whether the higher use mitigates mental health outcomes as we have put forward.

Finally, participants with hearing impairment were more likely to use social media to perform their work. Work can be challenging for people with hearing loss because of the difficulties in spoken communication. This can result in increased levels of fatigue, stress, loss of productivity, and job loss [49,50]. It would be of interest to study how adults with hearing impairment use social media to perform their work and whether social media use can help to mitigate these adverse effects. Consideration could also be given to using social media for audiological rehabilitation purposes, for instance, providing the option to consult vocational rehabilitation services through direct messaging, but this needs further investigation.

Strengths and Limitations

The strengths of this study include the large sample, its diversity in age, the comparison between adults with normal hearing and those with hearing impairment, the correction for confounders, and the functional measurement of hearing ability. As all measurements were performed through a questionnaire and a hearing test that ran in a web browser, the study was likely to

attract adults who are fairly comfortable with using the internet through a web browser. Nevertheless, the total sample seems representative of the general Dutch population in their use of technology and social media.

Our study includes several limitations. We used existing questionnaires that are used for marketing purposes. We were not able to find any documents showing validity of these questionnaires. Nevertheless, given that they are used in marketing research, we assume that they were developed according to accepted standards and hence valid, although a scientific report on their validity has not been published. We assume that the questions on use of technology and different types of apps were valid. The questions we devised on several aspects of social media use were not tested for validity either. This needs further investigation. In addition, all data were self-reported and could therefore suffer from recall bias, particularly the questions that included frequency of use, as well as from social desirability bias. The actual use of apps and social media on smart devices can be measured by individually tracking their use with software that collects these data and sends them to a research database [51], but this was not feasible in this study. Given that we did not ask about controversial topics with respect to using the internet and social media as well as the rather anonymous nature of filling out a web-based questionnaire, we believe that social desirability bias is of minimal importance in our study. Another limitation is the use of convenience sampling. The NL-SH sample is more highly educated than the general Dutch population, which could hamper generalization. However, the comparisons between adults with hearing impairment and those with normal hearing are still valid because they were performed within this sample and no influence of education level was found on the associations. All analyses were cross-sectional, which does not allow for inferences on causal relationships. Finally, we performed a very large number of statistical tests. The probability of a type 1 error, finding a statistically significant difference when in reality there is no difference, increases with multiple testing. We did not correct for this because we considered our study to be exploratory in nature. The results should be interpreted accordingly.

Conclusions

The potential for digitalized hearing health care is confirmed. Adults with impaired hearing are not less likely to use smart devices than their normal hearing peers. More research is needed into the needs and wants of adults with hearing impairment for the type of hearing health care solutions they seek.

Adults with hearing impairment agreed more with using social media to stay in touch with family members and friends than normal hearing adults, but this did not result in closer or more intense family ties or friendships. Research is needed on the type of social media that adults with hearing loss use to stay in touch with close family members and friends and to determine whether the higher use of social media mitigates mental health outcomes as we have put forward. Given that participants with hearing impairment are more likely than their normal hearing peers to use social media to perform their work, it would be of interest to study whether and how vocational rehabilitation services for workers with hearing impairment could be implemented on social media platforms as an alternative or supplementary to standard hearing health care. Adults with hearing impairment are less likely to make weekly use of social media apps on a smartphone, tablet, or smartwatch but not less likely to use social media on all types of internet-connected devices. This warrants further research on the types of social media platforms that adults with hearing impairment use and on the type of device on which they prefer to use social media.

Acknowledgments

The authors thank the participants of the Netherlands Longitudinal Study on Hearing. Funding for data collection of the 10-year measurement round came from the EMGO+ Institute for Health and Care Research, the Netherlands, and Sonova AG, Switzerland. This study was supported with monetary and nonmonetary means (ie, personnel time) by Sonova AG, Switzerland.

Conflicts of Interest

EU is affiliated with Sonova AG, the funder of this study.

Multimedia Appendix 1
Questions from the National Longitudinal Study on Hearing questionnaire used for this study.
[PDF File (Adobe PDF File), 517 KB - jmir_v23i12e27599_app1.pdf]

References

1. World Health Organization. Addressing the rising prevalence of hearing loss. Geneva: World Health Organisation; 2018:1-28.
2. Bainbridge KE, Wallhagen MI. Hearing loss in an aging American population: extent, impact, and management. Annu Rev Public Health 2014;35:139-152. [doi: 10.1146/annurev-publhealth-032013-182510] [Medline: 24641557]
3. Homans NC, Metselaar RM, Dingemanse JG, van der Schroeff MP, Brocaar MP, Wieringa MH, et al. Prevalence of age-related hearing loss, including sex differences, in older adults in a large cohort study. Laryngoscope 2017 Mar;127(3):725-730. [doi: 10.1002/lary.26150] [Medline: 27377351]
4. van der Meijden W, Panneman M, Kloet S, Blatter B, Homans N, Goedegebure A. Prevalentie van gehoorverlies in Nederland. VeiligheidNL, Kenniscentrum Letselpreventie.: VeiligheidNL; 2020. URL: https://www.veiligheid.nl/.ibmmodres/

domino/OpenAttachment/veiligheid/website.nsf/3F32154B9DC4F0B4C125868C0046F4B4/asset/Rapport%20prevalentie%20van%20gehoorverlies%20in%20Nederland.pdf [accessed 2021-11-17]
5. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media Society 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
6. Ageing Europe - statistics on social life and opinions. Eurostat. 2020. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_-_statistics_on_social_life_and_opinions#Education_and_digital_society_among_older_people [accessed 2021-07-15]
7. Paglialonga A, Nielsen A, Ingo E, Barr C, Laplante-Lévesque A. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng Online 2018 Jul 31;17(1):101 [FREE Full text] [doi: 10.1186/s12938-018-0531-3] [Medline: 30064497]
8. Paglialonga A, Tognola G, Pinciroli F. Apps for hearing healthcare. Stud Health Technol Inform 2015;210:666-668. [Medline: 25991233]
9. Meijerink JF, Pronk M, Lissenberg-Witte BI, Jansen V, Kramer SE. Effectiveness of a web-based SUpport PRogram (SUPR) for hearing aid users aged 50+: two-arm, cluster randomized controlled trial. J Med Internet Res 2020 Sep 22;22(9):e17927 [FREE Full text] [doi: 10.2196/17927] [Medline: 32960175]
10. Pronk M, Besser J, Smits C, Feenstra-Kikken V, van Beek H, Polleunis C, et al. Rationale, theoretical underpinnings, and design of HEAR-aware: providing adults with hearing loss with tailored support to self-manage their hearing problems via a smartphone app, as an alternative to hearing aids. Am J Audiol 2020 Sep 18;29(3S):648-660. [doi: 10.1044/2020_AJA-19-00079] [Medline: 32946257]
11. Saunders GH, Roughley A. Audiology in the time of COVID-19: practices and opinions of audiologists in the UK. Int J Audiol 2021 Apr;60(4):255-262. [doi: 10.1080/14992027.2020.1814432] [Medline: 32909474]
12. Swanepoel D, Hall J. Making audiology work during COVID-19 and beyond. Hearing J 2020;73(6):20-24. [doi: 10.1097/01.HJ.0000669852.90548.75]
13. Shukla A, Harper M, Pedersen E, Goman A, Suen JJ, Price C, et al. Hearing Loss, Loneliness, and Social Isolation: A Systematic Review. Otolaryngol Head Neck Surg 2020 May;162(5):622-633. [doi: 10.1177/0194599820910377] [Medline: 32151193]
14. Nachtegaal J, Smit JH, Smits C, Bezemer PD, van Beek JH, Festen JM, et al. The association between hearing status and psychosocial health before the age of 70 years: results from an internet-based national survey on hearing. Ear Hear 2009 Jun;30(3):302-312. [doi: 10.1097/AUD.0b013e31819c6e01] [Medline: 19322094]
15. Stam M, Smit JH, Twisk JW, Lemke U, Smits C, Festen JM, et al. Change in psychosocial health status over 5 years in relation to adults' hearing ability in noise. Ear Hear 2016;37(6):680-689. [doi: 10.1097/AUD.0000000000000332] [Medline: 27779518]
16. Blazer DG, Tucci DL. Hearing loss and psychiatric disorders: a review. Psychol Med 2019 Apr;49(6):891-897. [doi: 10.1017/S0033291718003409] [Medline: 30457063]
17. Chen E, Wood D, Ysseldyk R. Online social networking and mental health among older adults: a scoping review. Can J Aging 2021 Mar 19:1-14. [doi: 10.1017/s0714980821000040]
18. Gonsalves C, Pichora-Fuller MK. The effect of hearing loss and hearing aids on the use of information and communication technologies by community-living older adults. Can J Aging 2008;27(2):145-157. [doi: 10.3138/cja.27.2.145] [Medline: 18845510]
19. Henshaw H, Clark DP, Kang S, Ferguson MA. Computer skills and internet use in adults aged 50-74 years: influence of hearing difficulties. J Med Internet Res 2012;14(4):e113 [FREE Full text] [doi: 10.2196/jmir.2036] [Medline: 22954484]
20. Thorén ES, Oberg M, Wänström G, Andersson G, Lunner T. Internet access and use in adults with hearing loss. J Med Internet Res 2013 May 09;15(5):e91 [FREE Full text] [doi: 10.2196/jmir.2221] [Medline: 23659867]
21. Disabled users access to and use of communication devices and services - Research summary: all disabled people. Ofcom. 2019. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/132962/Research-summary-all-disabilities.pdf [accessed 2021-11-17]
22. Morris J, Sweatman W, Jones M. Smartphone use and activities by people with disabilities: user survey 2016. J Technol Persons Disab 2017;5:50-66 [FREE Full text]
23. Shoham S, Heber M. Characteristics of a virtual community for individuals who are d/deaf and hard of hearing. Am Ann Deaf 2012;157(3):251-263. [doi: 10.1353/aad.2012.1625] [Medline: 22978201]
24. Saxena RC, Lehmann AE, Hight AE, Darrow K, Remenschneider A, Kozin ED, et al. Social media utilization in the cochlear implant community. J Am Acad Audiol 2015 Feb;26(2):197-204 [FREE Full text] [doi: 10.3766/jaaa.26.2.8] [Medline: 25690778]
25. Choudhury M, Dinger Z, Fichera E. The utilization of social media in the hearing aid community. Am J Audiol 2017 Mar 01;26(1):1-9. [doi: 10.1044/2016_AJA-16-0044] [Medline: 28241238]
26. Crowson MG, Tucci DL, Kaylie D. Hearing loss on social media: who is winning hearts and minds? Laryngoscope 2018 Jun;128(6):1453-1461. [doi: 10.1002/lary.26902] [Medline: 28988431]

27. Bahng J, Lee CH. Topic modeling for analyzing patients' perceptions and concerns of hearing loss on social Q and A sites: incorporating patients' perspective. Int J Environ Res Public Health 2020 Aug 27;17(17):6209 [FREE Full text] [doi: 10.3390/ijerph17176209] [Medline: 32867035]
28. Kožuh I, Debevc M. The utilisation of social media among users with hearing loss: an analysis of Facebook communities. Univ Access Inf Soc 2019 Jun 18;19(3):541-555. [doi: 10.1007/s10209-019-00658-x]
29. National Hearing Test. URL: https://www.hoortest.nl/ [accessed 2021-01-29]
30. NL-SH study website for the general public and participants. National Longitudinal Study on Hearing. URL: https://www.hooronderzoek.nl/ [accessed 2021-01-29]
31. Kramer SE, Kapteyn TS, Festen JM. The self-reported handicapping effect of hearing disabilities. Audiology 1998 Jan;37(5):302-312. [doi: 10.3109/00206099809072984] [Medline: 9776207]
32. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. Int J Audiol 2004 Jan;43(1):15-28. [doi: 10.1080/14992020400050004] [Medline: 14974624]
33. Smits C, Merkus P, Houtgast T. How we do it: the Dutch functional hearing-screening tests by telephone and internet. Clin Otolaryngol 2006 Oct;31(5):436-440. [doi: 10.1111/j.1749-4486.2006.01195.x] [Medline: 17014457]
34. Arends-Tóth J. ICT-gebruik van huishoudens en personen [ICT use of households and persons]. ICT, kennis en economie. 2019. URL: https://longreads.cbs.nl/ict-kennis-en-economie-2019/ict-gebruik-van-huishoudens-en-personen/ [accessed 2020-12-08]
35. Smartphone penetration by age group (2012-18). Deloitte. 2018. URL: http://www.deloitte.co.uk/mobileuk2018/charts/uk-smartphone-penetration-by-age-group/ [accessed 2020-12-08]
36. Ledbury C, Woolhead L, Thomas F, Carpenter R, Stott H. Something for everyone. Why the growth of mobile apps is good news for brands. ipsos MORI. 2017. URL: https://www.ipsos.com/sites/default/files/2017-08/Google-mobile-apps-report-2017.pdf [accessed 2020-12-08]
37. Monitor uw eigen gezondheid met deze tien apps. Patientline Whitepaper. 2019. URL: https://patientline.nl/wp-content/uploads/2019/02/Whitepaper-top-10-gezondheidsapps.pdf [accessed 2021-11-17]
38. Van der Veer N, Boekee S, Hoekstra H, Peters O. Nationale Social Media Onderzoek 2018 [National Social Media Study 2018]. Newcom Research and Consultancy. 2018. URL: https://www.bindinc.nl/wp-content/uploads/2018/04/Newcom-Nationale-Social-Media-Onderzoek-2018-3.pdf [accessed 2021-11-17]
39. Blank G, Lutz C. Representativeness of social media in Great Britain: investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. Am Behav Sci 2017 Jun 29;61(7):741-756. [doi: 10.1177/0002764217717559]
40. Kožuh I, Debevc M. Challenges in social media use among deaf and hard of hearing people. In: Dey N, Babo R, Ashour A, Bhatnagar V, Bouhlel M, editors. Social Networks Science: Design, Implementation, Security, and Challenges. Cham: Springer; 2018:151-171.
41. Hoffman DL, Novak TP. Why do people use social media? Empirical findings and a new theoretical framework for social media goal pursuit. SSRN J 2020. [doi: 10.2139/ssrn.1989586]
42. Ngai EW, Tao SS, Moon KK. Social media research: theories, constructs, and conceptual frameworks. Int J Inform Manag 2015 Feb;35(1):33-44. [doi: 10.1016/j.ijinfomgt.2014.09.004]
43. Perugini ML, Solano AC. Normal and maladaptive personality traits as predictors of motives for social media use and its effects on well-being. Psychol Rep 2020 May 06;124(3):1070-1092. [doi: 10.1177/0033294120922495] [Medline: 32375534]
44. Heimlich R. Using social media to keep in touch. Pew Research Center. 2011. URL: https://www.pewresearch.org/fact-tank/2011/12/22/using-social-media-to-keep-in-touch/ [accessed 2020-12-08]
45. van Manen T. Onderzoek: het gebruik van internet en social media in Nederland [Research: the use of internet and social media in the Netherlands]. Marketingfacts. 2017. URL: https://www.marketingfacts.nl/berichten/onderzoek-het-gebruik-van-internet-en-social-media-in-nederland [accessed 2020-12-08]
46. Allen KA, Ryan T, Gray DL, McInerney DM, Waters L. Social media use and social connectedness in adolescents: the positives and the potential pitfalls. Aus Edu Develop Psychol 2020 Oct 19;31(1):18-31. [doi: 10.1017/edp.2014.2]
47. 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 19;46(2_suppl):69-80. [doi: 10.1177/1090198119863768] [Medline: 31742462]
48. Nielsen AC, Rotger-Griful S, Kanstrup AM, Laplante-Lévesque A. User-innovated eHealth solutions for service delivery to older persons with hearing impairment. Am J Audiol 2018 Nov 19;27(3S):403-416 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0009] [Medline: 30452745]
49. Nachtegaal J, Festen JM, Kramer SE. Hearing ability in working life and its relationship with sick leave and self-reported work productivity. Ear Hear 2012;33(1):94-103. [doi: 10.1097/AUD.0b013e318228033e] [Medline: 21826005]
50. Nachtegaal J, Kuik DJ, Anema JR, Goverts ST, Festen JM, Kramer SE. Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. Int J Audiol 2009;48(10):684-691. [doi: 10.1080/14992020902962421] [Medline: 19863354]
51. Araujo T, Wonneberger A, Neijens P, de Vreese C. How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. Commun Methods Measures 2017 Apr 27;11(3):173-190. [doi: 10.1080/19312458.2017.1317337]

Abbreviations

NHT: National Hearing Test
NL-SH: National Longitudinal Study on Hearing
OR: odds ratio
RQ: research question
SNR: signal-to-noise ratio
SRTn: speech-reception threshold in noise

Edited by R Kukafka; submitted 29.01.21; peer-reviewed by A Paglialonga, C Price, JR Bautista; comments to author 29.03.21; revised version received 03.08.21; accepted 22.09.21; published 20.12.21.

Please cite as:
van Wier MF, Urry E, Lissenberg-Witte BI, Kramer SE
A Comparison of the Use of Smart Devices, Apps, and Social Media Between Adults With and Without Hearing Impairment: Cross-sectional Web-Based Study
J Med Internet Res 2021;23(12):e27599
URL: https://www.jmir.org/2021/12/e27599
doi:10.2196/27599
PMID:34932013

Original Paper

EpiHacks, a Process for Technologists and Health Experts to Cocreate Optimal Solutions for Disease Prevention and Control: User-Centered Design Approach

Nomita Divi[1*], BSc, MSPH; Mark Smolinski[1*], MD, MPH

Ending Pandemics, San Francisco, CA, United States

[*]all authors contributed equally

Corresponding Author:
Nomita Divi, BSc, MSPH
Ending Pandemics
870 Market Street
Suite 528
San Francisco, CA, 94102
United States
Phone: 1 6173591733
Email: nomita@endingpandemics.org

Abstract

Background: Technology-based innovations that are created collaboratively by local technology specialists and health experts can optimize the addressing of priority needs for disease prevention and control. An EpiHack is a distinct, collaborative approach to developing solutions that combines the science of epidemiology with the format of a hackathon. Since 2013, a total of 12 EpiHacks have collectively brought together over 500 technology and health professionals from 29 countries.

Objective: We aimed to define the EpiHack process and summarize the impacts of the technology-based innovations that have been created through this approach.

Methods: The key components and timeline of an EpiHack were described in detail. The focus areas, outputs, and impacts of the twelve EpiHacks that were conducted between 2013 and 2021 were summarized.

Results: EpiHack solutions have served to improve surveillance for influenza, dengue, and mass gatherings, as well as laboratory sample tracking and One Health surveillance, in rural and urban communities. Several EpiHack tools were scaled during the COVID-19 pandemic to support local governments in conducting active surveillance. All tools were designed to be open source to allow for easy replication and adaptation by other governments or parties.

Conclusions: EpiHacks provide an efficient, flexible, and replicable new approach to generating relevant and timely innovations that are locally developed and owned, are scalable, and are sustainable.

(J Med Internet Res 2021;23(12):e34286) doi:10.2196/34286

KEYWORDS

epidemiology; public health; diagnostic; tool; disease surveillance; technology solution; innovative approaches to disease surveillance; One Health; surveillance; hack; innovation; expert; solution; prevention; control

Introduction

Technology can play an important role in the prevention and control of infectious disease. Historically, developing technical solutions has required substantial financial resources, human capacity, and time. All too often, innovative technologies that are introduced as pilot projects or research endeavors are terminated prematurely due to a lack of sustained funding or a lack of a postresearch plan. Third-party technologies that are created for governments are often discarded when vendors change or fail to create long-term plans for sustainability and scaling. In other cases, misalignment with current contexts and capabilities has often resulted in technologies being unsuccessful despite the best intentions. In short, technology solutions that are created *for* low- and middle-income countries to address local challenges, rather than those created *by* such countries, are quickly becoming a thing of the past.

Creating scalable and sustainable innovations for disease prevention and control requires local technologists to work

together with health experts to optimize the desired results when creating technology-based innovations. Conducting an EpiHack is one way to optimize such results. EpiHacks provide a unique process for combining the science of epidemiology with elements of a hackathon to cocreate solutions that address existing and future needs through a highly collaborative and intensive event.

Engaging local technologists and health experts from idea inception to prototype development can ensure that the appropriate contexts, such as language, culturally relevant imagery, and social norms, are duly considered when creating any solution to identified challenges. Additionally, innovative solutions that are generated through the use of existing infrastructures and widely available technologies are more likely to be widely adopted and affordable.

In an EpiHack, design thinking is the principal approach that is used to identify needs and guide the development of tools and systems to meet those needs. Design thinking is a process for solving problems by prioritizing end users' needs through observing how people interact with their environments and by using an iterative, hands-on approach to creating innovative solutions [1]. The focus of an EpiHack can be specifically on disease, situations, or improved data analysis and visualization. EEpiHack solutions have served to improve surveillance for influenza, dengue, and mass gatherings, as well as laboratory sample tracking and One Health surveillance, in rural and urban communities. As of January 2021, a total of 12 EpiHacks have been executed across 5 continents, resulting in open-source technologies positively impacting disease prevention and control. This paper defines the EpiHack process and summarizes the impacts of the technology-based innovations that have been created through this approach.

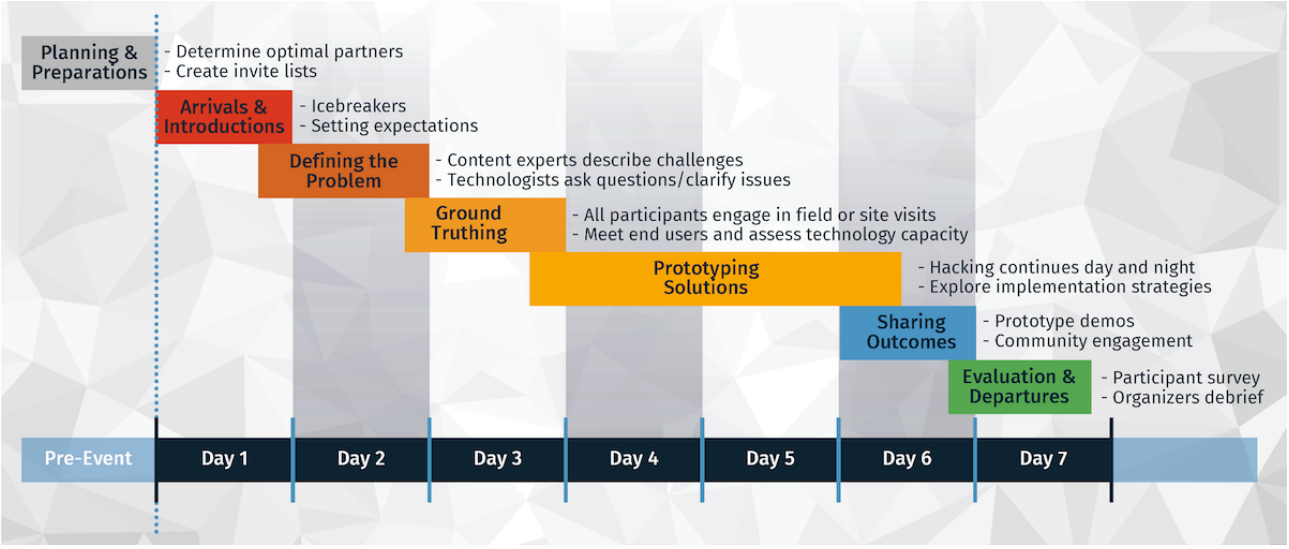
Methods

EpiHack elements and the timeline of an EpiHack were described. The twelve EpiHacks conducted between 2013 and 2021 were reviewed to document the focus areas, outputs, and impacts.

EpiHack Components and Timeline

Each EpiHack prioritized the needs of the host country or organization to solve real-world problems efficiently and effectively through a highly interactive and collaborative process. EpiHacks are immersive, residential programs in which developers and health experts work side by side on-site until the desired outcomes are achieved. EpiHacks generally occur over the period of 1 week but can be adapted according to the complexity of the problem(s) to be solved (Figure 1). In theory, it may be possible to host a web-based EpiHack, although this approach has not yet been tested. Participants do not compete for prizes or awards; rather, they collaborate to achieve common goals and maximize the desired impact.

Figure 1. EpiHack components and timeline. EpiHacks are ideally conducted over a 1-week period to provide optimum time for defining the problems, exploring solutions, and arriving at prototypes by the conclusion of the event.



Planning and Preparation

An EpiHack requires rigorous advanced planning, including securing a location for hosting an immersive, potentially weeklong event with extended hours each day. A good deal of time and effort goes into identifying participants, which include key stakeholders from the government, the private sector, and the community along with technologists and facilitators, who are critical to ensuring the successful outcomes of the event. It is imperative that various levels of government be brought into the planning process at the earliest stage, given their critical roles and responsibilities in national and local disease prevention and control.

The host organization is typically an in-country partner who is eager to improve local disease prevention and control capacities and fill existing gaps. Host organizations have included government agencies, universities, and not-for-profit organizations. The host must oversee all planning and the execution of the event and ensure that the right mix of health and technology participants will be in attendance. An appropriate balance of subject matter expertise and technology development

skills is crucial to inspiring unbridled creativity. A developer to health expert ratio of 3:1 is recommended.

Given that 3 in 4 emerging infectious diseases have their origins in animals, it is imperative that technology-enabled solutions for disease prevention and control engage both the human and animal sectors [2]. EpiHacks therefore require the active engagement of subject matter experts across the human and animal health sectors, including epidemiologists, veterinarians, and health care specialists. They also engage diverse end users, such as farmers, market workers, community health volunteers, district health officers, and lab technicians. EpiHacks also require the participation of a diverse set of local technology professionals who are familiar with front-end development (user interfaces and web design), back-end development (servers and databases), and data science elements (data management and analytics). Local technology expertise promotes mindfulness toward how proposed innovations can manifest into readily usable technologies.

Ensuring that strategic governmental partners, potential funders, and key community organizations are engaged in the planning of an EpiHack also improves the likelihood of successful outcomes. This also allows for the longer-term engagement of stakeholders from key implementing organizations, who are necessary for fully deploying the tools and systems that are created during an EpiHack.

In some cases, hosts may work with stakeholders in advance to frame the challenges that will be addressed during an EpiHack before the start of the event. In most instances however, the true nuances of these challenges come to light only after an EpiHack begins. Many aspects of an EpiHack are intentionally made to be unconfined to provide participants with the freedom to arrive at the most creative and appropriate solutions. Part of the initial planning includes arrangements for "field trips" or site visits to explore technical capabilities and to meet the potential end users of the innovations to be created during an EpiHack.

Most EpiHack attendees—both technology and health expert participants—reside in the host country. This ensures that participants are aware of the local contexts and challenges related to developing compelling prototypes that resonate with the targeted end users. Attracting and identifying the right local technology partners often require posting about the upcoming EpiHack on various media outlets and recruiting them from both the private and public sectors. Health expert attendees are generally invited based on their area of expertise or role within the government with regard to disease prevention and control.

Arrivals and Introductions

Given that many EpiHack participants do not know each other or have not attended a prior EpiHack, it is important to provide adequate time for getting acquainted through icebreakers, providing an introduction to the process of an EpiHack, and setting expectations for the days of the EpiHack. Often, the human and animal health experts, even if they are from the same country, do not know each other well and appreciate the opportunity to socialize before the intensive EpiHack process ensues. Building mutual trust and respect between the health experts and the technologists is essential, given the rapid pace of the event.

Defining the Problem

In most EpiHacks, participants begin their substantive work by defining the problem or clarifying the challenges to be addressed over the course of the week. This focus at the start provides the opportunity for content experts to explain existing challenges and allows for technology participants to ask questions to the health experts and vice versa. In cases where the problem to be addressed is determined through pre-event work, the EpiHack starts off with the host organization updating the participants on the desired goals.

Ground Truthing

Once the challenges have been defined, the participants travel to the "field" to visit relevant sites, so that they can better understand the needs of end users. During these visits, participants interact directly with multiple end users of the proposed technology solution(s) and explore their operating environments. This may include visits to government facilities, community health clinics, hospitals, or live animal markets to examine the existing workflows and current infrastructures with which the proposed tools will operate. These visits might entail engaging with farmers in rural villages or observing community health workers in their catchment areas to understand firsthand the challenges to be resolved. EpiHack solutions must be capable of integration and interoperation with existing systems and use the pervasive technologies in a given country or region. These solutions must also account for issues, such as those related to internet connectivity and ease of use.

Prototyping Solutions

An important part of an EpiHack is hacking, and this takes place over the several days of the so-called *hacking stage*. Participants work collaboratively to iterate on ideas, test concepts by developing mock-ups, and begin to create functional prototypes. During this time, technology experts work closely with their health counterparts to resolve issues related to design, testing, and strategies for implementation. Although hacking is the primary function of the technologists, the health experts use this time to further develop plans for pilot tests, end-user engagement, and the scaling of the solutions. Often, the participants divide themselves into smaller teams to concentrate on building a specific tool or to allow for intense focus on a particular issue.

All tools, prototypes, and systems that are developed in EpiHacks are open source, meaning that the software for original source codes is made freely available and may be modified. This allows others to replicate or adapt the tools for their own purposes. Participants are asked to sign an open-source agreement to this effect, waiving any intellectual property rights for the ideas or tools that are developed during EpiHack events. The designs, prototypes, and tools that are created during an EpiHack are typically further developed into fully deployed systems that serve the host country or organization following the event. In many cases, the collaborations among the technologists and health experts that occur during an EpiHack continue well after the week's end.

Sharing Outcomes

The penultimate event of the week is the showcasing of the solutions and prototypes developed during the EpiHack among the participants and key stakeholders. Community members are also invited to this event to broaden engagement and further facilitate the uptake of the solutions. Private sector participation can often help with devising strategies and creating necessary partnerships for financing and scaling tools. Such partnerships are especially useful when large-scale health emergencies require the rapid adoption and widespread use of solutions.

Evaluation and Departures

The work of an EpiHack does not end after participants share the outcomes. Often before leaving, participants make commitments to certain outcomes and align their roles with those of other stakeholders to ensure the completion and sustainability of the proposed solutions. Many times, participants create a mailing list or a WhatsApp group to update each other on the progress of tool development post-EpiHack and to share opportunities for further engagement. At the conclusion of each EpiHack event, participants complete a survey, and organizers hold an after-action debrief to evaluate the impact of the EpiHack and identify areas for improvement.

Research Ethics Statement

This study did not receive or require ethics approval, as it did not involve human or animal subjects.

Results

Since 2013, a total of 12 EpiHacks have been conducted, and each event resulted in 1 or more prototypes of tools or systems for enhanced disease prevention and control (Table 1).

EpiHack solutions have included national hotlines, One Health participatory surveillance systems, health monitoring tools for mass gatherings, and systems for mosquito vector control. EpiHack Phnom Penh, for example, laid the groundwork for the 115 Hotline—an interactive voice response system that allows Cambodians nationwide to report on suspected health threats in their community as well as access a menu of health education messages [3,4]. During the COVID-19 response, this free national hotline became the primary means for reporting COVID-19 suspect cases, initiating contact tracing efforts, and sharing COVID-19–related health information to the community. The hotline received an average of 18,000 calls per day at the peak of the COVID-19 pandemic and an average of 600 calls per day before the pandemic [5].

EpiHack Chiang Mai resulted in the Participatory One Health Disease Detection (PODD) system in Thailand, which uses smartphone apps and web applications to empower trained village volunteers to report unusual disease events in livestock, wild animals, and humans. By using this system, reports are triaged by PODD analysts and result in a response from local public health and livestock offices when warranted [6,7]. To date, over 19,000 volunteers have been trained to use the PODD system, and the project has expanded to approximately 400 subdistricts across 27 provinces in Thailand. In addition, community members used the PODD system to track the health of nonresidents who entered their village during the government-instituted COVID-19 lockdowns. Among 340 applications from 61 countries, the PODD system was the grand prize winner of the Trinity Challenge—an initiative that was created by a coalition of companies, foundations, and universities to recognize innovations in data-driven research and analytics for global health emergencies [8].

EpiHack Arusha participants created digital prototypes that resulted in the AfyaData mobile app for One Health surveillance in Tanzania. The system is used by trained volunteers in the Morogoro and Ngorongoro districts in Tanzania and allows for the real-time collection of data on human and animal health. AfyaData collects information at both the community and health facility levels and returns information on potential health intervention strategies to the reporters [9]. During the COVID-19 pandemic, AfyaData was shared with Mozambique for the cross-border tracking of COVID-19.

Due to the successful use of participatory surveillance for the first time at a mass gathering during the 2014 World Cup in Brazil, EpiHack Rio de Janeiro hosted a diverse set of international participants in 2015 to replicate this surveillance approach and enable the early identification of disease clusters during the Rio Olympic and Paralympic Games [10,11]. Members from India attended this EpiHack and incorporated their learnings from mass gathering disease surveillance during religious events in India [12]. The participatory surveillance tool in Brazil has since been adapted to monitor symptom reports from the public in Brazil and to capture data on COVID-19 symptom reporting in Brazil, Chile, and Peru.

Table 1. Summary of the EpiHacks conducted between 2013 and 2018.

Date	Location	Participants, n	Hosts	Focus area	Output and impact
August 2013	Phnom Penh, Cambodia	55	Royal University of Phnom Penh and Cambodia Ministry of Health	Participatory surveillance	The 115 national reporting hotline was created and deployed nationwide for disease reporting by the community and relaying health alerts. Averaging 600 calls per day, the hotline scaled during the COVID-19 pandemic to 18,000 calls per day and captured over 90% of the initial COVID-19 suspect cases [5].
March 2014	Chiang Mai, Thailand	49	Chiang Mai University College of Veterinary Medicine and Chiang Mai Provincial Public Health & Livestock Offices	One Health participatory surveillance	PODD[a] mobile apps and web-based applications were created. The PODD tool expanded from a single province and was adopted by >400 local governments across 27 provinces in Thailand with >19,000 trained volunteers [5]. During the COVID-19 pandemic, the PODD tool was adapted for tracking suspected COVID-19 cases.
June 2014	Lao People's Democratic Republic	62	Mekong Basin Disease Surveillance Network and Lao People's Democratic Republic Ministry of Health	Dengue surveillance	A dengue larvae survey tool was developed to allow for active community engagement in reducing the number of larvae breeding sites to improve dengue prevention and control.
December 2014	Arusha, Tanzania	74	Southern Africa Center for Infectious Disease Surveillance, Ministry of Health, and Ministry of Livestock and Fisheries	One Health participatory surveillance	The AfyaData One Health surveillance tool was developed, and it collected over 12,000 reports of abnormal health events in humans and animals over first 5 years of operation. The timeliness of local reporting to authorities improved from an average of 10 days to an average of 3 days [9]. During the COVID-19 pandemic, AfyaData was shared with Mozambique for cross-border reporting.
July 2015	Rio de Janeiro, Brazil	49	Brazil Ministry of Health	Mass gathering surveillance	The Guardians of Health surveillance tool was created, and it demonstrated the utility of participatory surveillance for monitoring health events during mass gatherings [10,11]. The tool has been used for COVID-19 surveillance among subpopulations in Brazil, Chile, and Peru.
September 2015	Minneapolis, United States of America	30	University of Minnesota Food Protection and Defense Institute and National Association of City and County Health Officials	Influenza surveillance	A Flu Near You (influenza-like illness symptom reporting system) data dashboard was created to improve data visualization, and it enhanced data analytics for use by local health departments.
April 2016	Saranda, Albania	41	Southeast European Center for Surveillance and Control of Infectious Diseases, Ministry of Health, Institute of Public Health, and Ministry of Animal Health	One Health surveillance systems	A digital data dashboard for One Health surveillance was built based on the learnings from the Thailand and Tanzania EpiHacks.
April 2016	Yangon, Myanmar	44	Myanmar Ministry of Health	Participatory surveillance	A hotline for citizen volunteers to report on health issues was developed for direct community reporting on health issues.
September 2016	Denver, United States of America	25	Council of State and Territorial Epidemiologists	Influenza surveillance	A Flu Near You data validation toolkit and data dashboard enhancements were developed for local health departments to better interpret Flu Near You data [13].
October 2017	Hanoi, Vietnam	31	Vietnam Ministry of Health General Department of Preventive Medicine	Participatory surveillance	A community outbreak reporting hotline was created by replicating the 115 Hotline model from Cambodia for use in Vietnam [14].

Date	Location	Participants, n	Hosts	Focus area	Output and impact
November 2017	Colombo, Sri Lanka	90	Colombo Municipal Council and Nanyang Technology University	Vector-borne disease surveillance	The Mo-Buzz dengue surveillance app was enhanced to improve the user interface, digitize paper forms, and improve data analytics and the visualization of breeding sites that were indicated as high-risk areas [15].
April 2018	Kampala, Uganda	44	Uganda Ministry of Health	Emergency Operations Center	A contact tracing prototype and laboratory sample tracking system were created.

[a]PODD: Participatory One Health Disease Detection.

EpiHacks have also contributed to building a community of practice around innovations in surveillance. In several instances, members who have attended an EpiHack apply their learnings to subsequent EpiHacks and new tool development. For example, health experts and developers from Tanzania attended EpiHack Chiang Mai (conducted in March 2014) to learn about the EpiHack process. Subsequently, they applied their learnings to the EpiHack in Tanzania (conducted in December 2014). Conversely, participants from EpiHack Chiang Mai attended the EpiHack in Tanzania to help guide the event. Furthermore, learnings from the Tanzania EpiHack informed the Albania EpiHack, which resulted in the development of the Albania One Health surveillance system.

EpiHack participants' evaluations uncovered key thematic areas, including the value of dedicating time to networking, the utility of the "field trips," and the desire for ongoing engagement with coparticipants. As noted in the evaluations, the provision of adequate time during each EpiHack for networking is prioritized. Additionally, up to 1 whole day is dedicated to the "field trip," so that all participants can experience the realities that are relevant to their solutions. Finally, given the interest of participants in continuing their engagement after EpiHack events, email listservs and WhatsApp groups are created by the organizers to facilitate ongoing communication among participants.

Discussion

EpiHacks provide an efficient and effective approach for generating relevant and timely innovations in disease prevention and control that are locally developed and owned. EpiHacks are well suited for tackling challenges that require creativity, consensus among stakeholders, and cross-sectoral collaboration. Political support and buy-ins from local leadership are foundational for a successful EpiHack. Although not all EpiHack hosts are government agencies, their participation is important for fostering the necessary buy-ins for the development of prototypes and sustainability.

EpiHacks offer opportunities for new collaborations and networking among participants at the national level and on a global scale. Most EpiHack participants are chosen by organizers from the host nation in keeping with a focus on local and sustainable solutions. Engaging EpiHack alumni, who understand the process due to their previous experiences, is highly beneficial. Those who are interested in hosting a future EpiHack might attend an event to experience this approach prior to hosting their own convening.

EpiHacks have been used primarily in the context of innovations for the prevention and control of infectious disease. This approach however could serve as a model for the advancement of solutions to other global health challenges. Web-based training on the EpiHack approach is free, and the health community is encouraged to consider this efficient method for advancing technology solutions to challenges in a variety of contexts [16]. The lessons gained from EpiHacks to date are that no community is too hard to reach, no country is too resource poor to innovate, and curiosity outshines fear around the globe when people are given the opportunity to create local solutions to pressing problems.

Acknowledgments

EpiHacks were conceived by Ending Pandemics—a San Francisco–based nonprofit—to support the locally led development of tools for rapidly detecting, verifying, reporting, and containing disease outbreaks across the One Health spectrum. Ending Pandemics would like to acknowledge everyone who participated in, facilitated, and organized EpiHacks, with a special thanks to the public health officials in Albania, Brazil, Cambodia, Laos, Myanmar, Sri Lanka, Tanzania, Thailand, Uganda, the United States of America, and Vietnam who engaged in this novel process for improving disease surveillance. We would also like to acknowledge the critical contributions of our former colleagues—Dr Jennifer Olsen, who facilitated several of the EpiHacks, and Adam Crawley, who assembled material for an early draft of the manuscript and Table 1. We are equally indebted to our partners from Opendream and InSTEDD (Innovative Support to Emergencies Diseases and Disasters) for designing and hosting numerous EpiHack events. Specifically, we would like to acknowledge Mr Patipat "Keng" Susumpow for coining the term *EpiHack* and Ms Prapapan "Bunny" Phetra and Mr Polawat "Pok" Phetra for their roles in developing the EpiHack web-based training resources and their invaluable contributions as EpiHack lead facilitators at events across the globe.

Authors' Contributions

ND was involved in the drafting, revision, and final review of the manuscript. MS conceptualized the EpiHack methodology. He was integrally involved in the drafting, revision, and final review of the manuscript.

Conflicts of Interest

None declared.

References

1. Tuttle G. What is design thinking and why is it important? WeWork. 2021 Feb 24. URL: https://www.wework.com/ideas/professional-development/creativity-culture/what-is-design-thinking [accessed 2021-09-20]
2. Woolhouse MEJ, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens. Emerg Infect Dis 2005 Dec;11(12):1842-1847 [FREE Full text] [doi: 10.3201/eid1112.050997] [Medline: 16485468]
3. Cambodia's 115 Hotline: Successful COVID-19 digital response. ICTworks. 2020 May 06. URL: https://www.ictworks.org/cambodia-115-hotline-digital-response/ [accessed 2020-05-11]
4. Case study: Cambodia's 115 Hotline: Expanding access to the reporting and detecting of disease outbreaks. Principles for Digital Development. URL: https://digitalprinciples.org/resource/case-study-cambodias-115-hotline-expanding-access-to-the-reporting-and-detecting-of-disease-outbreaks/ [accessed 2020-10-02]
5. Crawley A, Divi N. Surveillance technology in Thailand, Cambodia, and Tanzania: Case study from Ending Pandemics. Exemplars in Global Health. URL: https://www.exemplars.health/emerging-topics/epidemic-preparedness-and-response/surveillance-technology-ending-pandemics-case-study [accessed 2020-10-02]
6. Yano T, Phornwisetsirikun S, Susumpow P, Visrutaratna S, Chanachai K, Phetra P, et al. A participatory system for preventing pandemics of animal origins: Pilot study of the Participatory One Health Disease Detection (PODD) system. JMIR Public Health Surveill 2018 Mar 21;4(1):e25 [FREE Full text] [doi: 10.2196/publichealth.7375] [Medline: 29563079]
7. Pattamakaew S, Srikitjakarn L, Chatsowwong W, Bender JB. Factors associated with animal diseases and abnormal case reporting in disease surveillance by PODD Project's volunteer in Chiang Mai Province, Thailand. 2018 Jul Presented at: The 5th Food Saftey and Zoonoses Symposium for Asia Pacific; July 6-7, 2018; Chiang Mai, Thailand.
8. The Trinity Challenge winners. The Trinity Challenge. URL: https://thetrinitychallenge.org/awards/winners/ [accessed 2021-12-09]
9. Karimuribo ED, Mutagahywa E, Sindato C, Mboera L, Mwabukusi M, Njenga MK, et al. A smartphone app (AfyaData) for innovative One Health disease surveillance from community to national levels in Africa: Intervention in disease surveillance. JMIR Public Health Surveill 2017 Dec 18;3(4):e94 [FREE Full text] [doi: 10.2196/publichealth.7373] [Medline: 29254916]
10. Neto OL, Cruz O, Albuquerque J, de Sousa MN, Smolinski M, Cesse E, et al. Participatory surveillance based on crowdsourcing during the Rio 2016 Olympic Games using the Guardians of Health platform: Descriptive study. JMIR Public Health Surveill 2020 Apr 07;6(2):e16119 [FREE Full text] [doi: 10.2196/16119] [Medline: 32254042]
11. Mack H. Brazil health authorities launch app to map Zika during 2016 Olympics. MobiHealthNews. 2016 Jul 20. URL: https://www.mobihealthnews.com/content/brazil-health-authorities-launch-app-map-zika-during-2016-olympics [accessed 2019-01-14]
12. Diwan V, Hulth A, Manickam P, Balagurusamy VV, Agnihotri D, Parashar V, et al. Tablet-based participatory syndromic surveillance at Simhashta festival in India. Online J Public Health Inform 2018 May 22;10(1):1. [doi: 10.5210/ojphi.v10i1.8973]
13. Bakota E. Flu Near You, local data for local action. Council of State and Territorial Epidemiologists. 2017 Jun 26. URL: https://www.cste.org/blogpost/1084057/279336/Flu-Near-You-local-data-for-local-action [accessed 2021-12-09]
14. Vietnam 2018 national disease hotline launch. Innovative Support to Emergencies Diseases and Disasters. 2018 Nov 27. URL: https://instedd.org/blog/vietnam-2018-national-disease-hotline-launch/ [accessed 2021-12-09]
15. Lwin MO, Sheldenkar A, Panchapakesan C, Ng JS, Lau J, Jayasundar K, et al. Epihack Sri Lanka: development of a mobile surveillance tool for dengue fever. BMC Med Inform Decis Mak 2019 Jun 13;19(1):111 [FREE Full text] [doi: 10.1186/s12911-019-0829-5] [Medline: 31196073]
16. EpiHack. EpiHack. URL: http://epihack.org/ [accessed 2015-06-27]

Abbreviations

InSTEDD: Innovative Support to Emergencies Diseases and Disasters
PODD: Participatory One Health Disease Detection

Edited by G Eysenbach; submitted 15.10.21; peer-reviewed by T Yano; comments to author 05.11.21; revised version received 09.11.21; accepted 21.11.21; published 15.12.21.

Please cite as:
Divi N, Smolinski M
EpiHacks, a Process for Technologists and Health Experts to Cocreate Optimal Solutions for Disease Prevention and Control: User-Centered Design Approach
J Med Internet Res 2021;23(12):e34286
URL: https://www.jmir.org/2021/12/e34286
doi:10.2196/34286
PMID:34807832

Original Paper

Comparison of User-Oriented Information Services on the Websites of Large Hospitals in China and the United States: Cross-sectional Study

Yang Zhong[1*], MA, MBA; Wenjuan Tao[2*], PhD; Yanlin Yang[2], MS; Hao Wu[2], PhD; Weimin Li[3], MD, PhD; Jin Wen[2], PhD

[1]Academic Affairs Office, West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China

[2]Institute of Hospital Management, West China Hospital, Sichuan University, Chengdu, China

[3]President's Office, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Jin Wen, PhD
Institute of Hospital Management
West China Hospital
Sichuan University
No 37 Guoxue Xiang
Chengdu, 610041
China
Phone: 86 2885422671
Email: huaxiwenjin@163.com

Abstract

Background: Many people use the internet to access health care information to support health care decisions, and hospital websites can be the first point of contact to provide health care information services for consumers. However, little is known about the current information services provided by the websites of large Chinese hospitals.

Objective: The aim of this study is to evaluate and compare the information services of the websites of large hospitals in China and the United States. We hope that our findings will benefit hospital managers worldwide in providing service information on the web.

Methods: This study adopted a cross-sectional analytical approach to evaluate the websites of large hospitals in China and the United States in 2020. A total of 300 large hospitals were randomly selected, of which half were in China and half were in the United States. Based on the 7Ps marketing mix, we identified 39 items that represent typical hospital website information services, covering the following seven dimensions: product, price, place, propagation, people, process, and physical evidence.

Results: Most of the items (34/39, 87%) related to information services offered by hospital websites were less covered in China than in the United States; however, 5 items (appointments by a third-party platform, mobile payment, hospital value, hospital environment display, and physicians' profiles) had higher coverage in China. The average scores for hospital websites in China and the United States were 13.25 (SD 2.99) points and 23.16 (SD 2.76) points, respectively. Generally, high scores were given to the south areas of China and north areas of the United States.

Conclusions: Hospital websites in China lagged behind those in the United States with regard to information services offered. We recommend that hospital managers in China place more emphasis on the people, product, and propagation dimensions of the 7Ps marketing mix in the construction of information services on hospital websites. Through the comparison of the websites of large hospitals in China and the United States, our study findings can provide suggestions for forming standard hospital website construction guidelines worldwide.

(J Med Internet Res 2021;23(12):e27392) doi:10.2196/27392

KEYWORDS

hospital websites; internet; information services; marketing mix; 7Ps; health care information services; hospital management; hospitals; patient services; eHealth

Introduction

The internet has become ubiquitous in people's lives. As of October 2020, almost 5 billion people had direct access to the internet worldwide [1]. According to a report by the China Internet Network Information Center, as of March 2020, the penetration rate of internet users in China was 64.5%, which represents an increase of 4.9% when compared to this penetration rate in 2018 [2]. Most adults use the internet to access health care information to support health care decisions [3,4], and a hospital's home page can serve as the first point of contact to provide health care information services for consumers. Thus, in the internet era, the hospital website has become an important window connecting various departments of a hospital with patients and society, as well as a fairly new marketing tool to current and potential customers [5]. Customers' decision-making processes are often influenced by their perceptions and evaluations of a hospital's website [6]. It is important for hospital websites to meet their expectations by providing useful and accessible information [7]. Therefore, the evaluation of hospital websites has become inevitable.

The subject of hospital website evaluation has been heavily researched in different studies, such as an analysis of private hospital websites promoting medical tourism in India, Malaysia, and Thailand [8]; an evaluation of the quality of private hospital websites in Turkey; an evaluation of top academic hospitals for finding endocrine surgeons [7]; an evaluation of website information provided by pediatric surgery centers in Australia and New Zealand [4]; and a ranking of children's hospital websites in the United States [9]. However, previous studies have not fully examined the evaluation of websites of general or large hospitals. Various models or frameworks have been proposed to evaluate hospital websites, mainly focusing on the quality of the websites [10-12] or evaluating their accessibility [13]. Network engagement for hospitals is a skill that exists at the intersection of marketing and technical capability [6], and a well-designed hospital website that shows considerable service information could influence patients to take the first step into a facility. There is much less evidence of evaluating hospital websites from the perspective of information service.

Affected by the "Internet Plus Healthcare" strategy of the Chinese government (which proposes that internet technologies should be used to offer medical and public health services, promote family physician practices, improve drug supply and medical bill settlement, conduct medical education, and provide artificial intelligence services) [14], the overall construction level of hospital websites in China has improved significantly with time [15]. Furthermore, there is a substantial difference between past and modern hospital service concepts, especially in terms of offering user-oriented services and information. It is imperative to improve the information services provided by Chinese hospitals, especially those provided by large hospitals that offer intensive health care services. Therefore, it has become necessary to determine the current state of hospital websites in China. In addition, as shown in research that compared hospital websites from different countries [16], there is no study about the difference between hospital websites in China and other countries. The United States, where the internet emerged, was ranked third among countries with the most internet users after China, which had the highest number of internet users [17]. It would be interesting and meaningful to compare two powerful countries such as China and the United States, which have the largest economies in the world and represent a limited-income country and high-income country, respectively. There are 615,000 hospitals in the United States and 3,435,000 hospitals in China with a very large medical service market [18,19]. These hospitals could provide valuable information for global health care services when facing the increasing demand of the internet market.

This study aims to compare the information services of large hospital websites in China and the United States, based on the 7Ps marketing mix, and examine the problems that exist in Chinese hospital websites. We hope that our findings will benefit hospital managers not only in China and the United States but also around the world, especially in providing web-based service information.

Methods

Evaluation Framework Development

To the best of our knowledge, this study is the first to adopt the 7Ps marketing mix to form a logical evaluation framework for the sake of comparison, as this study focuses on the information services offered by hospital websites. The 7Ps marketing mix is a service marketing theory that was developed by Booms and Bitner [20] in 1981 by adding 3 new elements (people, process, and physical evidence) to the traditional marketing theory of 4Ps (product, price, place, and promotion). People, process, and physical evidence embody the characteristics of service marketing. Through a combination of production and consumption processes, this model allows customers to perceive high-quality service with the purpose of establishing, maintaining, and strengthening good long-term relationships with customers. According to the general definition and components of the 7Ps marketing mix, the corresponding definitions and components in relation to service information from hospital websites are presented in Multimedia Appendix 1. This evaluation framework consists of 7 dimensions and 39 items (components) for evaluating the service information of hospital websites. We consulted 10 hospital management or marketing experts to reach a consensus on determining and classifying the 39 items (components).

Sampling

A large cross-sectional analysis approach was adopted to evaluate hospital websites in China and the United States. We randomly selected 150 large Chinese hospitals and 150 large American hospitals (a total of 300 hospitals) to conduct a survey of their website information services. Because it is very difficult and time-consuming to extract information from websites, we decided to randomly take 3 and 5 hospitals in each state and province to obtain 150 samples from the United States and China, respectively, after consulting with a statistician. All eligible hospitals (over 500 beds) from each administrative region of the two countries were included initially, and then a random sampling procedure was conducted. We focused on large hospitals because they usually serve more patients on

average and their websites are visited by more users. Furthermore, large hospitals are more likely to represent the medical service capacity of a country. According to the current classification standards of Chinese hospitals, the definition of "large" refers to a Chinese hospital with at least 500 beds. Correspondingly, the same criterion was applied to the selection of American hospitals to map with the Chinese group. The number of beds for the selected hospitals in the United States was over 500 as well. The final sample of hospitals covered 31 provincial administrative regions in China (excluding Taiwan, Hong Kong, and Macao) and 45 states in the United States (excluding the states of Alaska, Montana, New Hampshire, Vermont, and Wyoming). Hospitals in Taiwan, Hong Kong, and Macao were excluded from the list of Chinese hospitals because their development and management systems are run differently from those in other regions of China for several historical reasons. Additionally, the population of these regions is too small to be included for explaining the general situation of China. We excluded 5 US states in our study because we failed to find hospitals with over 500 beds in these regions. Detailed information about the samples is provided in Multimedia Appendix 2.

Data Collection

From February to June 2020, two independent reviewers identified the websites of the sampled hospitals using the Baidu and Google search engines. Each website was comprehensively assessed based on the 39 items of the previously developed evaluation framework to check if the corresponding information was available.

Statistical Analysis

To create summarized scores of website service performance, the analytic method involved scoring the content according to 30 items, and each item was valued 1 point. The total score for each hospital website was reported in a range from 0 to 30, with a higher score on any given scale representing better comparative performance. The higher the score, the richer the information offered.

The 9 items that were not included in the scoring system were website accessibility, mobile payment, web account payment, appointment by telephone, appointment by official website, appointment by official app, appointment by a third-party platform, independent web page for hospital culture expressions, and patient-centered value. Website accessibility was not included in the scoring system because the information services and total scores for the websites could not be identified. Methods of making appointments and payments were excluded because there were no criteria showing which method was better. Independent web pages for hospital culture expressions and patient-centered value were not related to the availability of the service information offered but were related to how such information was offered by hospital websites; therefore, they were also left out of the scoring system.

Categorical variables were expressed as proportions and were tested using the chi-square test. Continuous variables were expressed as means and SDs and were tested by *t* test (if the sample observations followed a normal distribution and had homogeneity of variance) or Wilcoxon rank sum test (if the sample observations followed an abnormal distribution and had heterogeneity of variance). All analyses were conducted using Stata, version 12.0 (StataCorp LLC). All *P* values quoted below are 2-tailed and were considered significant when <.05.

Results

Of the sampled hospitals, a total of 143 Chinese and 150 American hospital websites were included, because 7 Chinese hospitals had no website on the internet or the site could not be accessed during our research.

All of the Chinese provinces/municipalities in this survey scored below 20 points, and more than 90% (29/31, 94%) scored in the range of 10 to 20 points. All of the studied American states scored above 20 points. As shown in Figure 1, there was a large gap between the two countries. It is interesting to note that the highest scores were generally given to the southern areas of China and northern areas of the United States. In terms of states or provinces/municipalities, the American hospital websites in Colorado, Delaware, and Rhode Island achieved the best scores (28 points); in China, the websites in Beijing, the capital city, obtained the highest score (17.2 points), followed closely by those in Zhejiang Province (17 points), Guangdong Province (15.43 points), and Shanghai Municipality (15.4 points). In the United States, hospitals in the state of California were rated the lowest (21.43 points); however, Sinkiang hospitals had the lowest score (5.4 points) in China.

The comparative results of 39 items representing the information services available on the websites between China and the United States are shown in Table 1. Overall, the information services offered by hospital websites in China were less covered than those offered in the United States, except for the following five items: appointment by a third-party platform, mobile payment, hospital value, hospital environment display (pictures/videos), and physicians' profiles. The item with the highest availability on websites in China and the United States was the medical services list (139/143, 97.2%) and appointment services (150/150, 100%), respectively. There were no information services about 2 items—web account payment and patient and family advisory council—on the Chinese hospital websites, and no appointment service by a third-party platform was available on American hospital websites. The data in Figure 2 are derived from the proportion of American hospital websites minus that of Chinese hospital websites for each information service item. Among the items in which Chinese hospital websites lagged behind, the five with the largest gaps were web account payment, patient privacy protection statement, volunteer services, COVID-19 information/policies, and social donation (Figure 2). The differences were all statistically significant (*P*<.001; Table 1).

Figure 1. Comparison of the scores of hospital website information services in China and the United States.

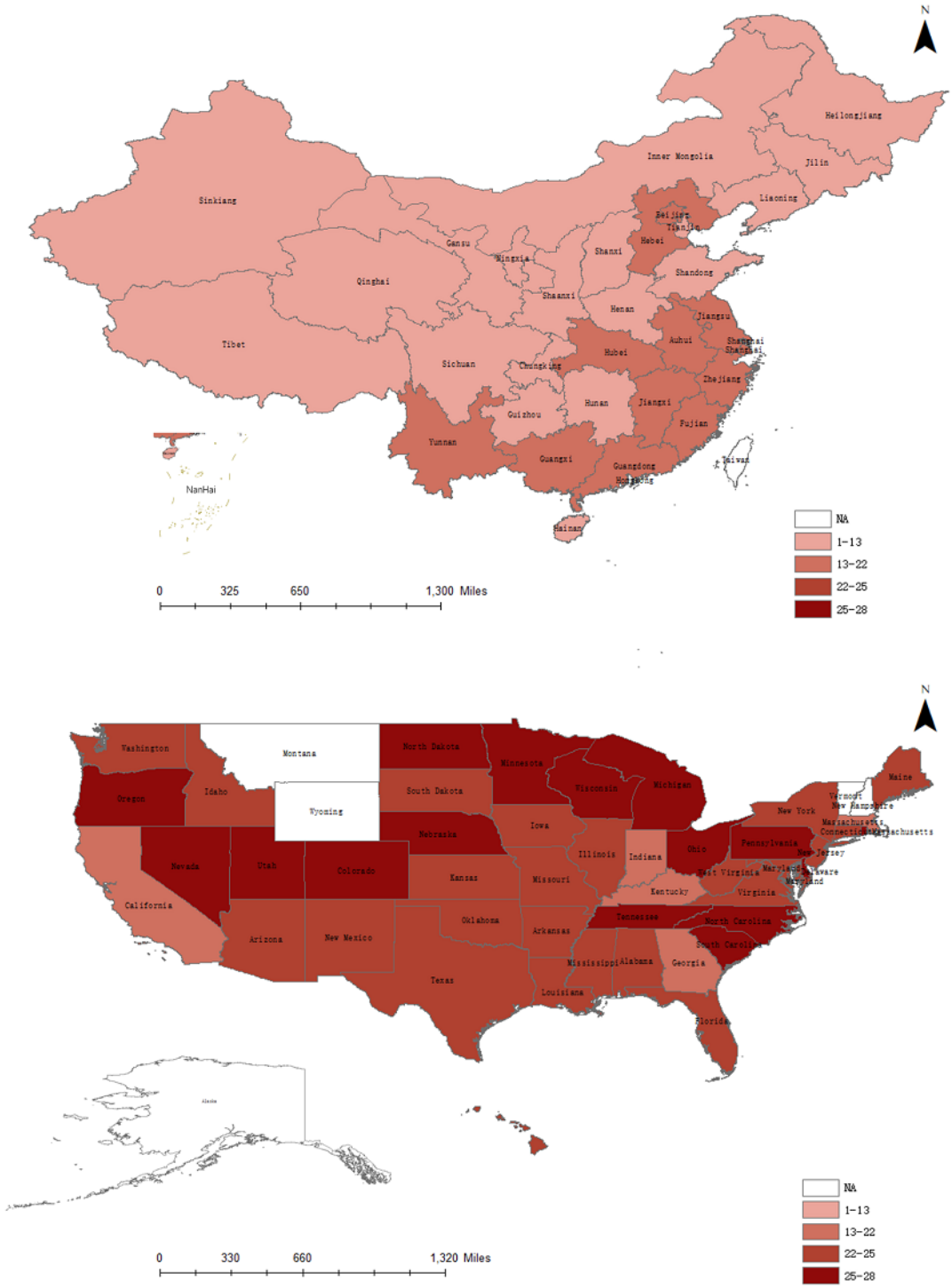


Table 1. Comparison of the available number of hospitals by information service items of hospital websites between China and the United States (N=300).

Dimension and items	China (n=143), n (%)	United States (n=150), n (%)	Chi-square (*df*)	*P* value
Product				
Medical services list	139 (97.2)	149 (99.3)	1.98 (1)	.16
Referral service	19 (13.3)	104 (69.3)	94.41 (1)	<.001
Online visit	82 (59)	132 (88.6)	33.94 (1)	<.001
Examination report query	94 (65.7)	134 (89.3)	23.62 (1)	<.001
Insurance services	103 (72)	132 (88)	11.76 (1)	.001
Living guide	8 (5.6)	95 (63.3)	107.06 (1)	<.001
Services for people with disabilities	4 (2.8)	95 (63.3)	119.92 (1)	<.001
Services for international patients	7 (4.9)	107 (71.3)	135.95 (1)	<.001
Price				
Pricing transparency	66 (46.2)	116 (77.9)	30.24 (1)	<.001
Online payment	95 (66.4)	142 (94.7)	37.75 (1)	<.001
Mobile payment	95 (66.4)	75 (50)	8.12 (1)	.004
Web account payment	0 (0)	141 (94.6)	N/A[a]	N/A
Place				
Web accessibility	143 (95.3)	150 (100)	7.17 (1)	.007
Traffic guide	121 (84.6)	147 (98)	16.8 (1)	<.001
Appointment services	136 (95.1)	150 (100)	7.52 (1)	.006
Appointment by telephone	43 (31.4)	143 (95.3)	128.38 (1)	<.001
Appointment by official website	40 (29.2)	115 (77.2)	66.2 (1)	<.001
Appointment by official app	20 (14.6)	81 (54.4)	49.4 (1)	<.001
Appointment by a third-party platform	119 (86.9)	0 (0)	N/A	N/A
Propagation				
Vision	64 (44.8)	102 (68)	16.11 (1)	<.001
Mission	34 (23.8)	142 (94.7)	153.38 (1)	<.001
Value	123 (86)	112 (74.7)	5.94 (1)	.02
Exclusive web page for hospital culture expressions	85 (59.4)	140 (93.3)	47.19 (1)	<.001
Patient-centered values	79 (55.2)	144 (96)	66.87 (1)	<.001
Health science information	119 (83.2)	137 (91.3)	4.37 (1)	.04
COVID-19 information/policies	13 (9.5)	135 (90.6)	188.08 (1)	<.001
Patient stories	14 (9.9)	97 (64.7)	92.99 (1)	<.001
Access to social media sites	109 (76.2)	140 (93.3)	16.79 (1)	<.001
People				
Physicians' profiles	136 (95.1)	139 (92.7)	0.75 (1)	.39
Patient and family advisory council	0 (0)	49 (32.7)	56.09 (1)	<.001
Patient privacy protection statement	16 (11.2)	143 (95.3)	208.85 (1)	<.001
Volunteer services	15 (10.5)	138 (92)	194.94 (1)	<.001
Social donation	13 (9.3)	131 (87.3)	176.45 (1)	<.001
Feedback channels for hospital and medical services	73 (51.1)	131 (89.7)	52.05 (1)	<.001
Feedback channels for website visit experience	5 (3.5)	13 (8.7)	3.39 (1)	.07
Process				

Dimension and items	China (n=143), n (%)	United States (n=150), n (%)	Chi-square (*df*)	*P* value
Classification of user-oriented interface	22 (15.4)	134 (89.3)	160.81 (1)	<.001
On-site search	101 (70.6)	143 (95.3)	32.08 (1)	<.001
Frequently asked questions	15 (10.5)	92 (61.3)	81.63 (1)	<.001
Physical evidence				
Hospital environment displays (pictures/videos）	138 (96.5)	134 (89.3)	5.66 (1)	.02

[a]N/A: not applicable.

Figure 2. Differences in proportions of information service items on American and Chinese hospital websites. The proportion difference equals the percentage of accessibility of an item on American hospital websites minus the percentage of accessibility of an item on Chinese hospital websites.

Table 2 indicates the scoring results of information services available on hospital websites in China and the United States. The average scores for the 143 hospital websites in China and the 150 hospital websites in the United States were 13.25 (SD 2.99) points and 23.16 (SD 2.76) points, respectively. The scores of American hospital websites on 6 dimensions (product, price, place, propagation, people, and process) were all higher than those of Chinese hospital websites, and the differences were all statistically significant (*P*<.001). For the physical evidence dimension, namely, hospital environment display (pictures/videos), the score of the Chinese hospital websites was higher than that of the hospitals in the United States, and the difference was statistically significant (*z*=–2.37; *P*=.02). In addition, it can be seen from Table 2 that the largest gap existed in the people dimension, with a difference of more than 2 times in the average scores.

Table 2. Scoring results based on the 7Ps marketing mix for Chinese and American hospital websites.

Criterion	China (n=143)		United States (n=150)		*z* score	*P* value
	Score, mean (SD)	Score, P_{50} (P_{25}, P_{75})[a]	Score, mean (SD)	Score, P_{50} (P_{25}, P_{75})[a]		
Product (n=8)	3.22 (1.13)	3 (2, 4)	5.31 (1.37)	6 (5, 6)	11.00	<.001
Price (n=2)	1.13 (0.68)	1 (1, 2)	1.72 (0.49)	2 (1, 2)	7.79	<.001
Place (n=2)	1.80 (0.45)	2 (2, 2)	1.98 (0.14)	2 (2, 2)	4.64	<.001
Propagation (n=7)	3.33 (1.23)	3 (2, 4)	5.77 (1.16)	6 (5, 7)	12.46	<.001
People (n=7)	1.85 (0.80)	5 (5, 6)	5.03 (0.88)	2 (1, 2)	14.67	<.001
Process (n=3)[b]	0.97 (0.69)	1 (1, 1)	2.46 (0.65)	3 (2, 3)	19.13	<.001
Physical evidence (n=1)	0.97 (0.18)	1 (1, 1)	0.89 (0.31)	1 (1, 1)	–2.37	.02
Total (n=30)	13.25 (2.99)	13 (11, 15)	23.16 (2.76)	23 (22, 25)	14.43	<.001

[a]P_{50}: 50th percentile; P_{25}: 25th percentile; and P_{75}: 75th percentile.

[a]Tested by a 2-tailed *t* test.

Discussion

Principal Findings

This is the first study to evaluate the website information services provided by large Chinese and American hospitals based on the 7Ps marketing mix. Overall, the hospital websites in the United States offer more information services compared to those in China, which suggests that US hospital managers consider this to be an important method of reaching their customers. According to the scores of the hospital websites, it can be clearly seen that the higher scores were mainly concentrated in the south and east regions of China, which have higher gross domestic products than those of other regions. This could be explained by the fact that these regions have more income and resources. It may also be the case that hospitals in these regions are more concerned about competition and have implemented a website as part of their strategy for attracting patients [21].

Specifically, the results suggest that large hospitals in China are largely lagging behind in offering user-oriented information services on their websites. Our findings revealed that the Chinese hospital websites focused primarily on basic services and information, such as the medical service list, appointment services, and physicians' profiles, rather than paying more attention to the people dimension and involving the participation of and communication with the public, including patients, patients' families, volunteers, social donors, and website users. Taking the item of the patient privacy protection statement as an example, the proportion difference between Chinese and American hospital websites was ranked first among the 30 comparable scoring items. This finding is consistent with a previous study, in which only 22% websites of leading Chinese general hospitals included a privacy and security policy or terms of use [22]. Similarly, a comparable study revealed that hospital websites in Kuwait focus primarily on promoting services provided by the hospital rather than on engaging and communicating with patients [23]; moreover, hospital websites in Italy function more as sources of information on admissions and services than as a means of communication between users and the hospital [24]. As for the product dimension, Chinese hospital websites are less thoughtful in caring about people with special needs, such as patients with disabilities or language communication barriers or people who require referral services or living guides.

For the propagation dimension, the function of disseminating public health information on websites is underused by large hospitals in China. At present, the COVID-19 pandemic is still active around the world; with the resumption of school and work, the Chinese government is implementing normalized epidemic prevention and control measures for a long-term fight against the disease. Hospitals in mainland China responded to the outbreak of COVID-19 positively by using information technology–enabled services [25]. Nonetheless, the number of hospital websites with COVID-19 prevention information and policies in the United States is approximately 10 times that of such websites in China. Additionally, the relatively low percentages we found for the items of vision, mission, and independent web pages for hospital culture expressions and patient-centered value on Chinese hospital websites indicate that there is significant potential for Chinese hospitals to promote their culture development.

Conversely, compared to the United States, our study shows that more Chinese hospital websites contain the following information services: appointment by a third-party platform, mobile payment, physicians' profiles, hospital value, and environment display.

Two factors can be mainly attributed to the differences in appointments by third-party platforms and mobile payments between the two countries. First, the payment method for medical expenses in a country is closely related to the country's medical insurance system. In the United States, large hospitals are often tied to (large) employers, such as Kaiser Permanente. Patients do not have the choice to "shop" for hospitals and may not wish to do so. This is also true of payment. Employees are often insured "for free" in the United States, as their employer provides them with insurance. People who are unemployed and have a low income are not insured, and they may have no money and no possibility to make mobile payments. In China, more

than 95% of people are covered by basic health insurance [26]. Chinese people are usually free to make choices when seeking care and making payments.

Second, with the continuous development of mobile information technology, some basic functions performed by hospital websites in the past have been gradually transferred to mobile terminals, and the proportion of mobile medical services used in the process of patient medical treatment is increasing in China [27]. Chinese people, especially the younger generation, are more accustomed to using phones to make mobile payments—usually the consultation fee, which is a relatively small amount—to immediately confirm their appointments on the web. Mobile payment in China has become a life habit for Chinese people; this is not only reflected in the payment of medical expenses but also in all aspects of life because of its convenience and popularity. Chinese hospitals, under the influence of the "Internet Plus Healthcare" strategy, which is the application of internet technology in the medical industry and has been promoted by the Chinese government since its initiation in 2015, are also willing to cooperate with some large-scale application platforms, such as WeChat and Alipay, to implement these functions conveniently.

Therefore, Chinese users can easily obtain what they need without visiting hospital websites. In contrast, the relatively high rate of appointments on official websites and web account payments by American hospital websites has firmly locked American users into using them, and these websites have become the main source for Americans to obtain health information and services. With regard to the hospital values and environment display, as well as physicians' profiles, we found in this study that incorporating hospital values, pictures of the hospital environment, and physicians' profiles into the hospital's introduction and home pages has almost become a fixed practice in most hospitals in China, which may partly explain the higher percentages of these services on Chinese hospital websites.

Strengths and Weaknesses of This Study

By referring to the 7Ps marketing mix, we examined the websites' information services in terms of the aspects of product, price, place, propagation, people, process, and physical evidence. To the best of our knowledge, no previous study has built evaluation items and thoroughly examined the information services of hospital websites based on the 7Ps marketing mix. In addition, this is the largest cross-sectional survey to date, assessing and comparing 150 hospital websites in China and 150 hospital websites in the United States.

This study may have several limitations that are important to note. On the one hand, we aimed to study the information services provided by hospital websites, which are different from traditional services; therefore, the classical 7Ps marketing mix may not be fully applicable to this study. However, we have adapted it to the websites' information service features to perform a rational evaluation. On the other hand, we simply assigned 1 point to each item to build the scoring system for evaluating the information services of Chinese and American hospital websites; this approach may be unreasonable, as it does not take into account the weight of each item. Finally, there were 119 private (79.3%) and 31 public (20.7%) hospitals among the sample of 150 American hospitals. All 150 Chinese hospitals in the sample were public. We did not take the matter of public and private hospitals into account; moreover, the differences in social, political, legal, cultural, and other environmental factors between China and the United States were not considered, which could have affected the results.

Potential Implications for Hospital Leaders

For hospital leaders, the findings of our study may have some potential implications based on the analysis of the product, price, place, propagation, people, process, and physical evidence dimensions of the 7Ps marketing mix.

Product Dimensions

It is necessary for hospital leaders to establish a user-friendly hospital website based on users' actual needs. A study by Hakim and Deswindi [28] showed that the functional aspects of hospital websites are the most significant dimension because customers want to obtain in-depth information about a hospital's organization, facilities, and list of services. Overall, it is necessary to reduce or categorize the information that has little relevance for website users; unfortunately, this problem is commonly found on the home pages of Chinese hospital websites. For example, Chinese hospital websites usually provide medical service information in the form of listings of departments; in contrast, most American hospital websites provide more detailed information about medical services, such as diseases, symptoms, departments, surgical operations, and other related services. This information is more helpful to web users who know little about their diseases or are not familiar with the related medical knowledge. For information services, it is important to strengthen the function of guidance rather than focusing on display only.

Price and Place Dimensions

With the development of mobile information technology, some important information services, such as payment and appointment methods, have been gradually transferred from websites to mobile terminals (eg, hospital apps and WeChat) in China. Therefore, customers no longer find it necessary to access a website to obtain required health information services. Mobile information technology has brought great convenience to people's lives, and its use has become an inevitable trend. Compared to hospital websites in China, it was found that in the United States, a better balance is reached between the use of mobile devices and official websites. How to take advantage of mobile technology while also building a website that is valuable to visit has become an important issue that is worthy of consideration by Chinese hospital leaders.

Propagation Dimension

Culture development in an organization is a complicated and long-term process; however, one important measure is to clearly establish the mission of the organization, and this is especially urgent for Chinese hospital leaders given our findings. Furthermore, Chinese hospitals should make full use of the propagation function of their website platforms, whether to provide information on how to fight the pandemic or to disseminate public health knowledge. Another notable fact is that American hospital websites often use patient stories to build

a positive image of their hospitals. This can readily enhance a customer's feelings of trust toward a hospital. However, Chinese hospitals usually propagate information about the hospitals from the perspectives of medical staff skills or medical specialty strengths. As a result, because of information asymmetry between providers and patients, it is inevitable that certain obstacles in understanding will be created for ordinary users and will keep them at a distance.

People Dimension

The largest difference between Chinese and American hospitals lies in the people dimension, in which Chinese hospitals require much improvement to reach the level of US hospitals. According to our findings, Chinese hospitals should enhance their interactions on their websites with users and attach sufficient importance to social participation from individuals and other organizations. In detail, Chinese hospitals need to work more on feedback regarding hospital services and website visit experiences, develop volunteer services and social donation channels on their websites, show more respect to patients by emphasizing patient privacy protection statements, and attempt to learn more about patient and family advisory councils to determine if these organizations could work at Chinese hospitals.

Process Dimension

In the context of this research, the process dimension refers to the information or functions that are provided by a website to help users obtain the required information services. We investigated the situation of interface classification by users, site searches, and frequently asked questions pages of hospital websites between China and the United States, and we found that all of them were less covered on Chinese hospital websites. Therefore, more attention should be paid to the process dimension, as it directly reflects user-oriented ideas and is beneficial for improving website browsing experiences.

Physical Evidence Dimension

Vivid pictures or videos can be used to make websites more appealing to potential customers, so that they will spend more time on the websites. Although a few more Chinese hospitals displayed pictures or videos of the hospital environment according to our research, it was found that hospital websites in the United States as a whole are more aesthetic in terms of their visual design, which is worthy of learning by Chinese hospital administrators.

Unanswered Questions and Future Research

In this paper, we studied the availability of information services on Chinese and American hospital websites; however, we did not investigate which factors could affect the performance of such websites. It would be interesting to know if there are certain correlations among some additional factors, such as the economic development level and population size of different regions in China and the United States, development level of informatization, and competition of medical services. This is worthy of analysis and discussion in further studies.

Conclusions

Overall, hospital websites in China lagged behind their counterparts in the United States in providing user-oriented services and information. Most Chinese hospital websites focused primarily on basic services and information rather than involving the participation of and communication with the public. Additionally, culture development and the dissemination of public health knowledge need to be strengthened on Chinese hospital websites. Notably, we found that the transfer of basic services and information to mobile terminals has reduced the number of visits to hospital websites and may create obstacles to their further development. Hospital administrators can use the recommendations with respect to the product, price, place, propagation, people, process, and physical evidence dimensions to improve their hospitals' websites. On the basis of the comparison of the large hospital websites in China and the United States, it is critical to develop standard website construction guidelines for hospitals worldwide.

Acknowledgments

Funding was received from the Special Funds for COVID-19 Prevention and Control of West China Hospital of Sichuan University (grant HX-2019-nCoV-023).

Authors' Contributions

YZ and WT wrote the first draft of the manuscript, collected and analyzed the data, produced the tables and figures, and interpreted the results. YY contributed to the literature search. HW helped produce the figures. JW and WL critically revised the manuscript and provided overall guidance. All authors approved the final submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
7Ps marketing mix definition and components in relation to information services on hospital websites.
[DOCX File , 19 KB - jmir_v23i12e27392_app1.docx]

Multimedia Appendix 2
List of sample hospitals in China and the United States.
[DOCX File , 68 KB - jmir_v23i12e27392_app2.docx]

References

1. Internet World Stats: Usage and population statistics. Internet World Stats. URL: https://internetworldstats.com/stats.htm [accessed 2020-12-14]
2. The 45th China Statistical Report on Internet Development. China Internet Network Information Center. 2020. URL: http://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm [accessed 2020-11-12]
3. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
4. Meiyappan V, Little TA, Jackson P. Evaluation of website information provided by paediatric surgery centres in Australia and New Zealand. ANZ J Surg 2019 Sep;89(9):1119-1121. [doi: 10.1111/ans.15012] [Medline: 30685891]
5. Revere L, Robinson Jr L. How healthcare organizations use the Internet to market quality achievements. J Healthc Manag 2010;55(1):39-49; discussion 49-50. [Medline: 20210072]
6. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
7. Gentry ZL, Ananthasekar S, Yeatts M, Chen H, Dream S. Can patients find an endocrine surgeon? How hospital websites hide the expertise of these medical professionals. Am J Surg 2021 Jan;221(1):101-105. [doi: 10.1016/j.amjsurg.2020.06.016] [Medline: 32622508]
8. Moghavvemi S, Ormond M, Musa G, Isa CRM, Thirumoorthi T, Bin Mustapha MZ, et al. Connecting with prospective medical tourists online: A cross-sectional analysis of private hospital websites promoting medical tourism in India, Malaysia and Thailand. Tour Manag 2017 Feb;58:154-163. [doi: 10.1016/j.tourman.2016.10.010]
9. Huerta TR, Walker DM, Ford EW. An evaluation and ranking of children's hospital websites in the United States. J Med Internet Res 2016 Aug 22;18(8):e228 [FREE Full text] [doi: 10.2196/jmir.5799] [Medline: 27549074]
10. Farrahi R, Gilasi H, Khademi S, Chopannejad S. Towards a comprehensive quality evaluation model for hospital websites. Acta Inform Med 2018 Dec;26(4):274-279 [FREE Full text] [doi: 10.5455/aim.2018.26.274-279] [Medline: 30692713]
11. Aladwani AM, Palvia PC. Developing and validating an instrument for measuring user-perceived web quality. Information & Management 2002 May;39(6):467-476. [doi: 10.1016/s0378-7206(01)00113-6]
12. Jeddi FR, Gilasi H, Khademi S. Evaluation models and criteria of the quality of hospital websites: a systematic review study. Electron Physician 2017 Feb 25;9(2):3786-3793 [FREE Full text] [doi: 10.19082/3786] [Medline: 28465807]
13. Acosta-Vargas P, Acosta T, Luján-Mora S. Framework for accessibility evaluation of hospital websites. 2018 Presented at: 2018 International Conference on eDemocracy & eGovernment (ICEDEG); April 4-6, 2018; Ambato, Ecuador. [doi: 10.1109/ICEDEG.2018.8372368]
14. China's cabinet issues guideline on 'Internet Plus healthcare'. National Health Commission of the People's Republic of China. 2018. URL: http://en.nhc.gov.cn/2018-04/28/c_73906.htm [accessed 2021-04-03]
15. Opinions of the General Office of the State Council on Promoting the Development of "Internet plus Health". General Office of the State Council. 2018. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2020-11-28]
16. Llinás G, Rodríguez-Iñesta D, Mira JJ, Lorenzo S, Aibar C. A comparison of websites from Spanish, American and British hospitals. Methods Inf Med 2008;47(2):124-130. [Medline: 18338083]
17. Pensworth L. 2019 Internet Statistics, Trends and Data. DailyWireless. URL: https://dailywireless.org/internet/usage-statistics/#Internet_Usage_Worldwide [accessed 2020-12-21]
18. Fast facts on U.S. Hospitals, 2021. American Hospital Association. URL: https://www.aha.org/statistics/fast-facts-us-hospitals [accessed 2020-12-10]
19. 2019 Statistical Bulletin of China's Health Development. National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/guihuaxxs/s10748/202006/ebfe31f24cc145b198dd730603ec4442.shtml [accessed 2020-12-10]
20. Booms BH, Bitner MJ. Marketing strategies and organization structures for service firms. In: Donnelly JH, George WR, editors. Marketing of Services. Chicago, USA: American Marketing Association; 1981:47-51.
21. Norem J, Moen MA. The websites of Norwegian hospitals: do they meet national guidelines and patient's expectations? J Telemed Telecare 2004;10(5):272-276. [doi: 10.1258/1357633042026260] [Medline: 15532132]
22. Liu X, Bao Z, Liu H, Wang Z. The quality and characteristics of leading general hospitals' websites in China. J Med Syst 2011 Dec;35(6):1553-1562. [doi: 10.1007/s10916-010-9432-0] [Medline: 20703762]
23. Alhuwail D, AlMeraj Z, Boujarwah F. Evaluating hospital websites in Kuwait to improve consumer engagement and access to health information: a cross-sectional analytical study. BMC Med Inform Decis Mak 2018 Sep 24;18(1):82 [FREE Full text] [doi: 10.1186/s12911-018-0660-4] [Medline: 30249244]

24. Maifredi G, Orizio G, Bressanelli M, Domenighini S, Gasparotti C, Perini E, et al. Italian hospitals on the web: a cross-sectional analysis of official websites. BMC Med Inform Decis Mak 2010 Apr 01;10:17 [FREE Full text] [doi: 10.1186/1472-6947-10-17] [Medline: 20359366]
25. Yan A, Zou Y, Mirchandani DA. How hospitals in mainland China responded to the outbreak of COVID-19 using information technology-enabled services: An analysis of hospital news webpages. J Am Med Inform Assoc 2020 Jul 01;27(7):991-999 [FREE Full text] [doi: 10.1093/jamia/ocaa064] [Medline: 32311036]
26. Tao W, Zeng Z, Dang H, Li P, Chuong L, Yue D, et al. Towards universal health coverage: achievements and challenges of 10 years of healthcare reform in China. BMJ Glob Health 2020 Mar 19;5(3):e002087 [FREE Full text] [doi: 10.1136/bmjgh-2019-002087] [Medline: 32257401]
27. Bingzhong J, Ying Y, Jiaqian M, Zhonglian H, Caisheng H, Zhongmin R. Construction status and thinking on hospital network service system based on investigation. Journal of Medical Informatics 2015;36(5):68-71. [doi: 10.3969/j.issn.1673-6036.2015.05.014]
28. Hakim L, Deswindi L. Assessing the effects of e-servicescape on customer intention: A study on the hospital websites in South Jakarta. Procedia Soc Behav Sci 2015 Jan 20;169:227-239. [doi: 10.1016/j.sbspro.2015.01.306]

Edited by R Kukafka; submitted 24.01.21; peer-reviewed by R van Poelgeest, C Reis, A Hidki; comments to author 16.03.21; revised version received 10.04.21; accepted 08.11.21; published 29.12.21.

Please cite as:
Zhong Y, Tao W, Yang Y, Wu H, Li W, Wen J
Comparison of User-Oriented Information Services on the Websites of Large Hospitals in China and the United States: Cross-sectional Study
J Med Internet Res 2021;23(12):e27392
URL: https://www.jmir.org/2021/12/e27392
doi:10.2196/27392
PMID:34964717

Original Paper

mUzima Mobile Electronic Health Record (EHR) System: Development and Implementation at Scale

Martin Chieng Were[1], MD, MSc; Simon Savai[2], BSc; Benard Mokaya[2], BSc; Samuel Mbugua[2], MSc; Nyoman Ribeka[3], MSc; Preetam Cholli[1], MD; Ada Yeung[1]

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Institute of Biomedical Informatics, Moi University, Eldoret, Kenya

[3]Digital Impact Alliance, Washington, DC, United States

Corresponding Author:
Martin Chieng Were, MD, MSc
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Avenue, Suite 750
Nashville, TN, 37203
United States
Phone: 1 615 322 9374
Email: martin.c.were@vumc.org

Abstract

Background: The predominant implementation paradigm of electronic health record (EHR) systems in low- and middle-income countries (LMICs) relies on standalone system installations at facilities. This implementation approach exacerbates the digital divide, with facilities in areas with inadequate electrical and network infrastructure often left behind. Mobile health (mHealth) technologies have been implemented to extend the reach of digital health, but these systems largely add to the problem of siloed patient data, with few seamlessly interoperating with the EHR systems that are now scaled nationally in many LMICs. Robust mHealth applications that effectively extend EHR systems are needed to improve access, improve quality of care, and ameliorate the digital divide.

Objective: We report on the development and scaled implementation of *mUzima*, an mHealth extension of the most broadly deployed EHR system in LMICs (OpenMRS).

Methods: The "Guidelines for reporting of health interventions using mobile phones: mobile (mHealth) evidence reporting assessment (mERA)" checklist was employed to report on the *mUzima* application. The World Health Organization (WHO) Principles for Digital Development framework was used as a secondary reference framework. Details of *mUzima*'s architecture, core features, functionalities, and its implementation status are provided to highlight elements that can be adapted in other systems.

Results: *mUzima* is an open-source, highly configurable Android application with robust features including offline management, deduplication, relationship management, security, cohort management, and error resolution, among many others. *mUzima* allows providers with lower-end Android smartphones (version 4.4 and above) who work remotely to access historical patient data, collect new data, view media, leverage decision support, conduct store-and-forward teleconsultation, and geolocate clients. The application is supported by an active community of developers and users, with feature priorities vetted by the community. *mUzima* has been implemented nationally in Kenya, is widely used in Rwanda, and is gaining scale in Uganda and Mozambique. It is disease-agnostic, with current use cases in HIV, cancer, chronic disease, and COVID-19 management, among other conditions. *mUzima* meets all WHO's Principles of Digital Development, and its scaled implementation success has led to its recognition as a digital global public good and its listing in the WHO Digital Health Atlas.

Conclusions: Greater emphasis should be placed on mHealth applications that robustly extend reach of EHR systems within resource-limited settings, as opposed to siloed mHealth applications. This is particularly important given that health information exchange infrastructure is yet to mature in many LMICs. The *mUzima* application demonstrates how this can be done at scale, as evidenced by its adoption across multiple countries and for numerous care domains.

(J Med Internet Res 2021;23(12):e26381) doi:10.2196/26381

KEYWORDS

mobile health; electronic medical records; developing countries; digital divide; digital health; global health

Introduction

Background

Low- and middle-income countries (LMICs) have, over the last decade, seen an exponential increase in the adoption of digital health solutions. Among the systems being actively implemented in these settings are electronic health record (EHR) systems. These are deployed to largely replace or supplement existing paper-based records, with the aims of improving quality of patient care and supporting the monitoring and evaluation of programs [1-3]. Several LMICs have gone beyond initial pilot EHR system implementations to large-scale rollout of these systems in government-run public facilities. Countries like Kenya, Uganda, Nigeria, and Mozambique now run nationally endorsed EHR systems that are deployed in hundreds to thousands of public health facilities across each country [4-7].

In most LMICs, national-level EHR system initiatives have largely been driven by the need to support HIV care and treatment, with significant funding coming from donor organizations such as the U.S. President's Emergency Plan for AIDS Relief. EHR systems targeting HIV care largely focus on the HIV care continuum, which emphasizes (1) finding patients who are HIV-positive (through active screening approaches), (2) linking HIV-positive patients to care, (3) ensuring that patients are on appropriate treatment, and (4) actively following patients to retain them in care [8-10]. The continuum of care paradigm is widely applicable for numerous other chronic diseases and is employed within other EHR systems to support longitudinal care in LMICs.

Approaches and Gaps in HIV EHR System Implementations Within LMICs

There are several core functionalities needed within EHR systems to support longitudinal care, key among them being the abilities to register patients, review historical patient information, and collect new data on patients. Features such as computerized decision support, order entry capabilities, and electronic prescribing are often incorporated at varying levels [11]. In many settings, retrospective entry of data is still employed, though efforts are underway to increase use of EHR systems at the point of care [12]. Point-of-care EHR systems still face the challenges of inadequate infrastructure, unreliable system uptime, cost, busy care settings, and provider discomfort with real-time entry of data while caring for patients.

The most widely employed modality of EHR system implementation in LMICs involves standalone EHR system implementations at individual facilities [13]. This implementation model entails installation of a local server, local area network, and end user terminals. The model relies on a dependable electrical supply and readily available information technology personnel [14]. Unfortunately, these infrastructure and personnel requirements are prohibitive in many LMIC settings, particularly in remote areas—exacerbating the "digital divide." Facilities in areas with limited power, internet connectivity, and technical support are thus less likely to implement EHR systems [15]. Further, population health programs and community-based care services occurring outside of care facilities, such home-based HIV screening, testing, and patient tracing, are often poorly supported when access to EHR systems is limited to facilities.

To address some of the challenges and gaps with tethered EHRS that can only be accessed from within facilities, mobile health (mHealth) solutions are increasingly being adopted in LMICs [16,17]. Ideally, these mHealth systems should exchange data with the predominant EHR system, but a majority do not [18]. Instead, most mHealth systems often collect and transmit data to their own independent repositories that are separate from facility-based EHR systems—adding to the problem of siloed information for patients spread across multiple systems [19]. In the absence of interoperability with facility-based EHR systems, mHealth solutions cannot receive and display comprehensive historical patient and treatment information that are stored in the separate EHR systems, adversely affecting quality of care. Even when mHealth systems are able to share data with facility-based EHR systems, there is often lack of robust deduplication approaches for patient data, causing further challenges. Other common challenges observed with attempts to synergize mHealth solutions with facility-based EHR systems include (1) a lack of robust mechanisms to generate subsets of patients from EHR systems to be availed on mobile devices that have limited data storage capacities; (2) inefficient data synchronization, with mHealth applications requiring a complete wiping of existing data prior to new updates, resulting in increased expenses from use of paid internet data known as data bundles; (3) suboptimal mechanisms for updating or adding new forms, with some mHealth solutions requiring a new version of the application to be installed every time new forms are deployed; and (4) inability to handle forgotten log-in credentials that are aligned with EHR system credentials when providers are off-site.

With renewed emphasis on reaching the "last digital mile" while ensuring health information exchange with existing EHR systems, mHealth solutions are needed that can seamlessly interoperate with existing EHR systems in LMICs. This exchange is needed even in settings where health information exchange infrastructure is yet to mature [20]. In this paper, we present one such mobile application, *mUzima* [21], a UNICEF-recognized digital global public good [22], as a demonstration of a successful extension of existing national-level EHR systems in several LMICs, with the aim of increasing access and reach of digital technologies for providers and patients. We report on *mUzima*'s features and functionality, guided by the mHealth Evidence Reporting and Assessment (mERA) guidelines [23] and referencing the World Health Organization (WHO) Principles for Digital Development framework [24] where relevant. We also provide real-world examples of how the *mUzima* application has been scaled to support care across several countries and disease domains.

Methods

The "Guidelines for reporting of health interventions using mobile phones: mobile (mHealth) evidence reporting assessment (mERA)" checklist was developed by the WHO mHealth Technical Evidence Review group to improve comprehensiveness and standardization of reporting of mHealth interventions [23]. The guidelines ensure that the reporting covers the (1) content of the mHealth intervention, (2) context within which the mHealth intervention is implemented, and (3) technical features of the intervention. These 3 components are encapsulated in the *mERA* checklist of 16 elements, namely (1) infrastructure (population level), (2) technology platform, (3) interoperability/health information system (HIS) context, (4) intervention delivery, (5) intervention content, (6) usability/content testing, (7) user feedback, (8) access of individual participants, (9) cost assessment, (10) adoption inputs/program entry, (11) limitations for delivery at scale, (12) contextual adaptability, (13) replicability, (14) data security, (15) compliance with national guidelines or regulatory statutes, and (16) fidelity of the intervention.

This paper describes in detail each element of the mERA checklist as relevant to the *mUzima* application. Special emphasis is placed on the technology platform (item #2) to highlight choices made and features of the application that can serve as a technical reference for readers. As the *mUzima* application is developed for a global context, additional considerations need to be given to its applicability in LMICs. To this end, we also integrate components of the Principles of Digital Development by the WHO, using this framework to further elucidate the various items in the mERA checklist [24]. The Principles are "designed to help integrate best practices into technology-enabled programs...and are derived from lessons learned through the use of information and communication technologies (ICTs) in development projects" [24]. The 9 principles include (1) design with the user; (2) understand the existing ecosystem; (3) design for scale; (4) build for sustainability; (5) be data driven; (6) use open standards, open data, open source, and open innovation; (7) reuse and improve; (8) address privacy and security; and (9) be collaborative. Reference to these principles extends the relevance of the reported work to the larger LMIC and global context.

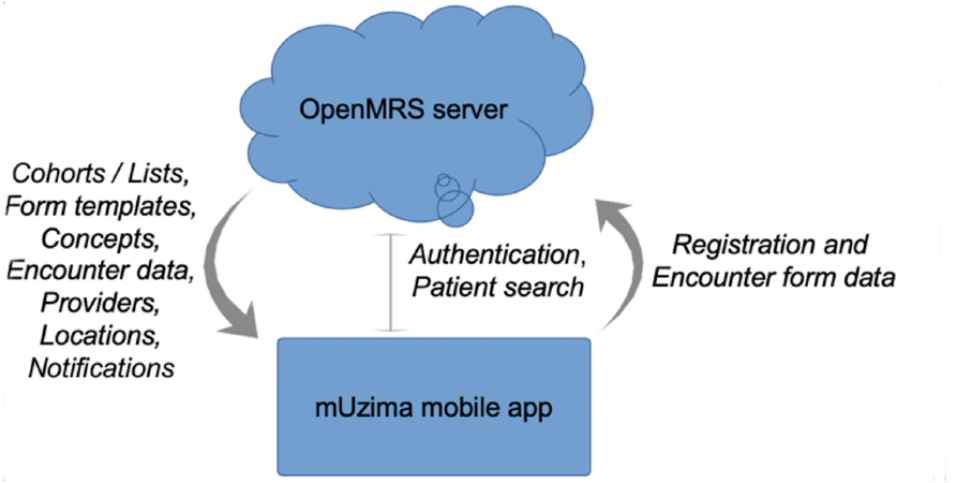
Results

'*Uzima*' is a Swahili word that means "life"—hence *mUzima's* slogan "Mobile for Life" [21]. It is a provider-facing mHealth application for use by health care workers for direct patient care. *mUzima* has particular relevance to providers working outside of care facilities and those with unreliable connection to the EHR system. In the following sections, we provide details of the *mUzima* application in line with 16 mERA checklist items.

Item 1: Infrastructure

mUzima is a robust and adaptable Android-based mHealth platform that can seamlessly interoperate with the OpenMRS EHR system (Figure 1) [25,26]. The OpenMRS EHR system was chosen for the mHealth extension, as it is the most widely endorsed EHR system for national use by Ministries of Health (MoHs) across numerous LMICs. OpenMRS is currently in use in over 40 countries, with national deployments for HIV and tuberculosis care in Kenya, Mozambique, Nigeria, and Uganda, among others [27]. The hardware requirements for *mUzima* include an Android smartphone device and an OpenMRS EHR system instance. In Africa, over 85% of smartphones use the Android operating system [28]. *mUzima* was developed to work with Android versions 4.4 and above, as low-end Android smartphones predominate in LMICs. *mUzima* has robust offline functionality to address limited internet connectivity in many LMICs, allowing for health providers using *mUzima* to care for patients even when disconnected from the EHR system server. These infrastructure considerations align with the Digital Development Principles of "Understanding the Existing Ecosystem" by recognizing that power and internet connectivity challenges exist in the LMIC settings where *mUzima* is intended for use and that low-end Android smartphones predominate in these settings.

Figure 1. Interaction between *mUzima* and the OpenMRS electronic health record (EHR) system.

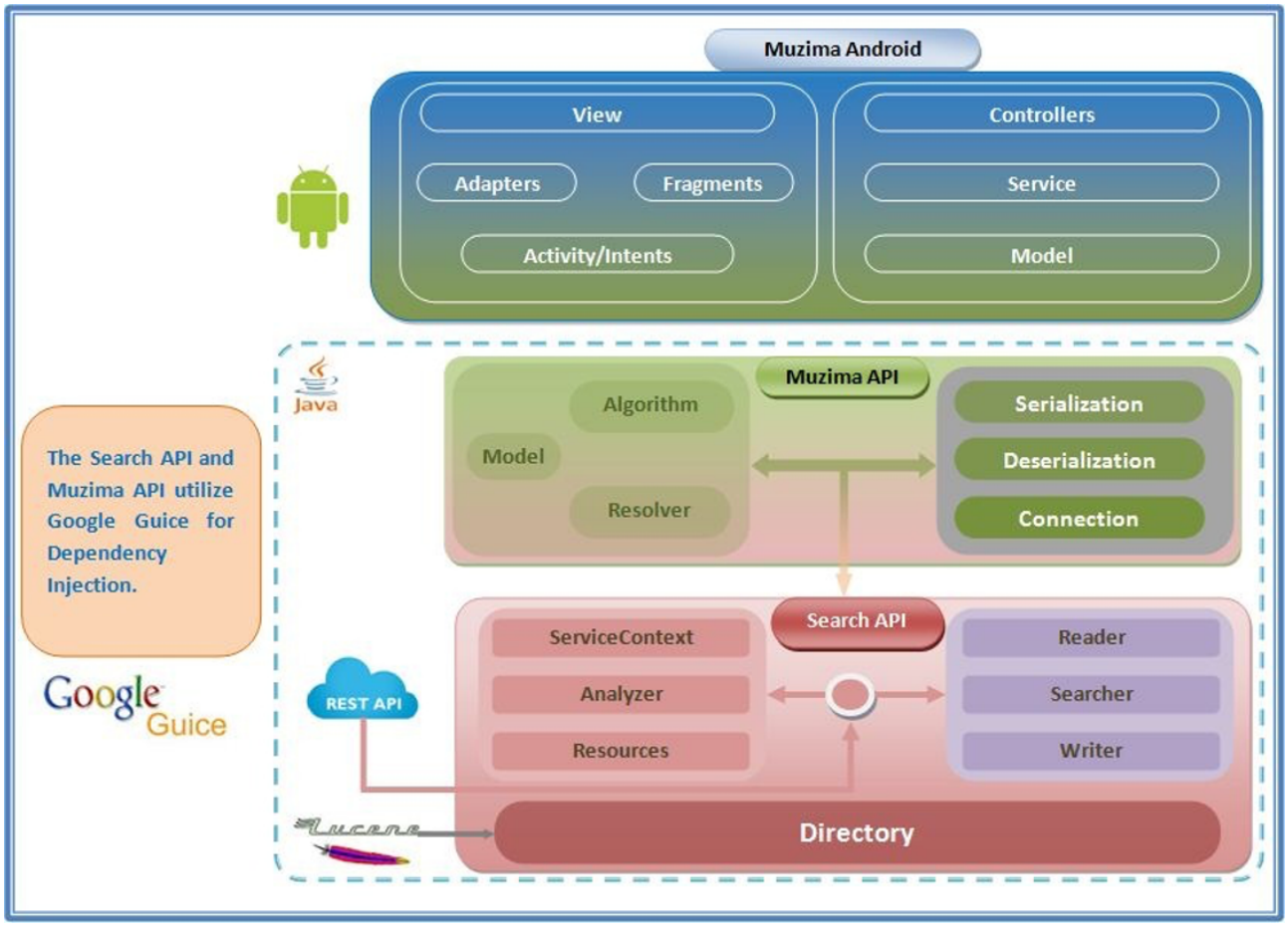


Item 2: Technology Platform

Overview

mUzima was developed as an open-source application under the Mozilla Public License 2.0 license [29]. The application has a modular architecture that lends it a strength of simplicity while ensuring full functionalities of the installed modules. *mUzima* uses a stacked framework consisting largely of 3 layers (Figure 2), namely the (1) Search Application Programming Interface (API), (2) *mUzima* API, and (3) *mUzima* Android.

Figure 2. Architecture of the *mUzima* mobile health (mHealth) application. API: application programming interface; REST: representational state transfer.



The Search API provides encrypted data storage based on a robust and flexible data model. This API utilizes Lucene, a Java-based, open-source, searching engine library, for indexing and searching [30]. Within the Search API are the service layer and a dependency injection layer using Guice [31] and the Indexer document repository [32]. Connection between *mUzima* and the EHR system server occurs using a representational state transfer (REST) API [33].

The *mUzima* API includes the business logic layer of the platform, exposing the underlying Lucene repository to a robust front-end application (*mUzima* Android) while hiding the complexity of the underlying repository to the developer. This layer can be easily used to expose the *mUzima* data model to any system integrating with *mUzima*.

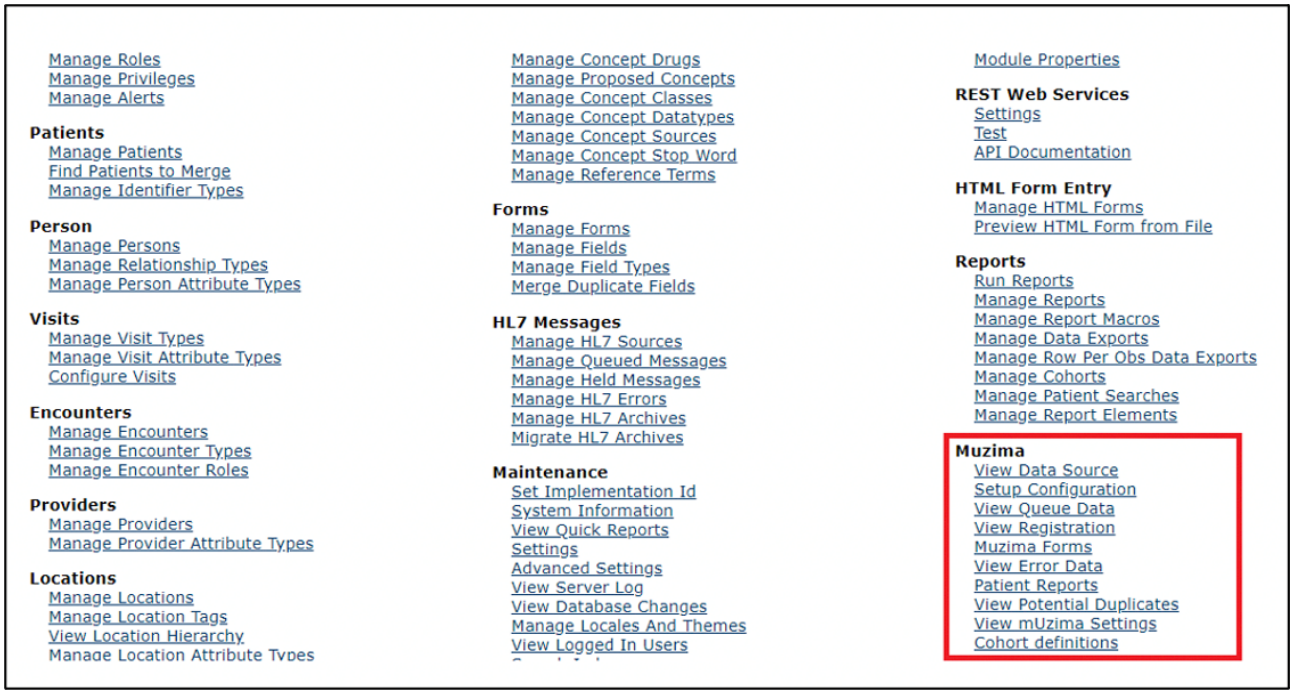
mUzima Android is the Android application package (APK) with which users interact, oblivious of the other 2 underlying stack layers. *mUzima* Android supports user interface activities and uses the underlying *mUzima* API to store and retrieve data resources from the phone's local repository and to upload and download data from the remote EHR system server. The APK is structured into the following main components: (1) models comprised of objects that encapsulate data and logic for resources specific to Android. Some of the objects extend *mUzima* API models and add logic specific to usage in the Android application and (2) controllers for each type of resource. The controllers act as coordinators within the Android app and between the Android app and *mUzima* API. Controllers are commonly used to request various actions related to objects, such as sending to storage, retrieval from storage, searching, counting, uploading, and downloading; (3) services, which are components used to handle requests for repetitive and potentially long-running operations; and (4) views, which are the implementation of the *mUzima* user interface.

mUzima Server-Side Module

mUzima has a server-side module created within the OpenMRS EHR system that has REST endpoints enabling linkage of OpenMRS instances to the *mUzima* mobile application (Figure 3 and Multimedia Appendix 1). Key features of the *mUzima* server-side module include (1) the set-up configuration and (2) error resolution mechanisms. The server-side setup configuration features are used to define care programs. To this end, the setup configuration has provisions for defining patient lists or cohorts,

forms, historical data, providers, and locations to be included in each care program (Multimedia Appendix 1). The server side also offers the ability to define multiple settings that will be applied onto all *mUzima* instances on mobile devices, removing the labor-intensive need of having to configure individual settings within each mobile device. The error resolution mechanism on the server side allows nondevelopers to manage forms submitted to the server that are identified as having errors (Multimedia Appendix 2). Errors are typically encountered when information on clients registered offline are synchronized onto the server and are found to be duplicates once patient-matching is run on the server. The ability to resolve errors on submitted data without the need for programming expertise or access to the server's backend makes *mUzima* highly usable and scalable in remote facilities, which often lack personnel to offer advanced technical support. The error resolution module also handles queuing and processing of data submitted into OpenMRS from *mUzima*.

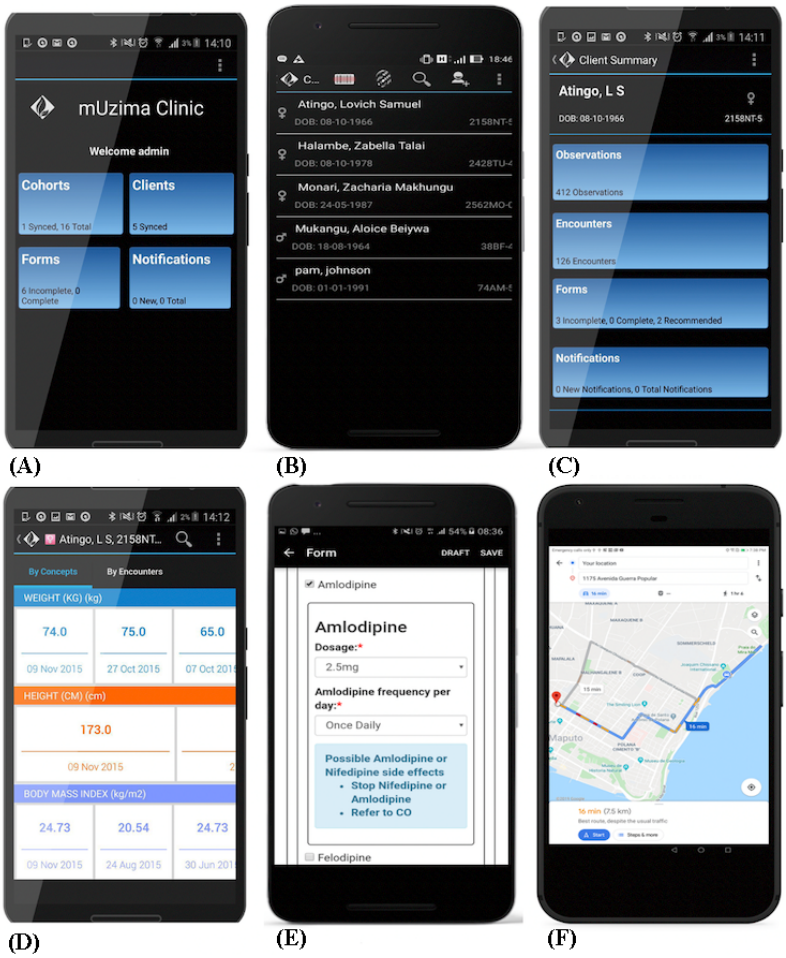
Figure 3. Components of the *mUzima* server-side module.



mUzima Mobile Application

Access to the *mUzima* mobile application requires a username and password, which are initially authenticated against the EHR system. This access times out automatically if the user is idle for a predefined period of time. Once logged in, users have access to the *mUzima* mHealth application landing page, which can be defaulted to display key menu options of cohorts, forms, clients, and notifications (Figure 4A). Alternatively, the landing page can be configured to display lists of clients available on the device (Figure 4B). *mUzima* users are able to find clients by scrolling through the client list, by searching via names and IDs, or by using the barcode scanning feature. Partial matching of names is available against both the local Lucene-based database that provides fast search capabilities [28] and the linked OpenMRS instance via its RESTful API. Further details of specific *mUzima* features are outlined in the Item 4 and Item 5 sections that follow.

Figure 4. *mUzima* features: (A) menu options, (B) client list, (C) options under client, (D) historical data, (E) form entry with decision support, (F) geolocation.



Item 3: Interoperability and Health Information System (HIS) Context

mUzima serves as an HIS for collecting primary health data, which are then exchanged with the associated EHR system. Health information exchange between *mUzima* and OpenMRS is achieved through use of the same concept dictionary terms, locations, providers, and patient identifiers that are common between the 2 systems. The REST API is used to connect *mUzima* to OpenMRS for bidirectional data exchange (Figure 1) [33]. This multifaceted approach for direct system-to-system exchange has been adopted to overcome the lagging implementation of robust health information exchange mediators and mechanisms in many LMICs. To facilitate future integration and health information exchange with other systems, several fast health care interoperability resources (FHIR) such as patient, provider, and observation FHIR are being developed for *mUzima* [34].

Item 4: Intervention Delivery

The WHO, in its "Recommendations on digital interventions for health system strengthening," strongly advocates that "recommended interventions should be accessible via mobile devices at a minimum" [35]. *mUzima* is primarily for use by health providers through smartphones that are often owned by the care program. Access to *mUzima* is on-demand when the provider logs in with the relevant credentials. The providers use *mUzima* to access cohorts and historical data on patients under their care, as well as to capture new data on patients. Significant time and effort were expended to ensure that *mUzima* can auto-update changes in cohort membership, forms, and historical data elements (ie, managing delta), without the need to re-download data that had previously been downloaded onto the mobile device. This allows for seamless update of content to the mobile device through *mUzima's* synchronization mechanism with OpenMRS.

Item 5: Intervention Content

Over half of the recommended digital interventions by the WHO pertain to mobile-based EHR solutions and include "digital tracking of clients' health status combined with decision support; targeted client communication; digital provision of educational and training content to health workers; provider-to-provider telemedicine; and client-to-provider telemedicine" [35]. *mUzima* satisfies all of the WHO mobile-based EHR system recommendations. Core features currently available within the *mUzima* application are described in the following sections.

Cohort Management

Given limitations on storage size within mobile devices, approaches are needed that limit the amount of data downloaded onto devices. This means downloading only the relevant subset of patients and their corresponding data to the mobile device. *mUzima* features the ability to define cohorts of patients using

SQL. Cohorts can be as simple as a list of individuals with a particular disease (eg, diabetes) or can be more complicated, such as "diabetic patients with kidney disease plus poorly controlled blood sugar and who are noncompliant with their medications." *mUzima* has functionality that allows for membership within a cohort to remain "static" (ie, never change after the initial run) or to be dynamic (allowing for automatic addition or removal of members over time, based on whether they still meet the cohort criteria).

Programs

A program within *mUzima* is used to consolidate all features in the mHealth application that are needed to take care of a group of patients. Programs can be disease-specific or based on cohorts. *mUzima* provides the ability to configure all details for programs using the set-up configuration feature on the server side (see the *mUzima* Server-Side Module section). Programs contain (1) lists of clients from one or more cohorts, (2) details of relevant historical data elements to be downloaded onto the device for these clients, (3) program-specific forms to be completed for clients, (4) an optional list of locations and providers that serve the program, and (5) settings that can be predefined for end users. Program functionality enables *mUzima* to be highly configurable and disease-agnostic, allowing the application to serve multiple clinical scenarios and domains. Further, programs integrate workflow features, such as sequential filling of forms, to support care processes. The ability to easily configure elements for programs ensures needed optimization of mobile device setup and also reduces the cost and time of data transmission and downloads.

Historical Data

Through its connection with an EHR system, *mUzima* has the ability to download a subset of historical data on patients that are predefined as part of a program. Beyond program-defined historical data to be downloaded, users of the *mUzima* application have the option to include additional historical data or remove any or all of the predefined historical data, based on their individual preferences. The historical data downloaded through *mUzima* can come from any source that saves data to the EHR system, including laboratory and pharmacy systems, ensuring that providers using *mUzima* have the complete clinical picture of a patient's data. This contrasts with numerous mHealth applications that only display historical data originally collected through that application and not from any other sources or those that are used as simple data collection tools, with no ability to display any historical information. Within *mUzima*, the historical data can be viewed by data type, date, or clinical encounter type (Figure 4D).

Data Capture Mechanisms

mUzima allows flexibility in data collection, with forms created to capture elements based on the same concept dictionary terms and other patient attributes (eg, demographics) shared with the EHR system. HTML5 is used to describe each form, enabling form development with little programming experience. *mUzima* forms can accommodate all data types, such as text, numeric, float, Boolean, and dates, and can support numerous media types. The application can also capture individual data elements (eg, individual test results), without relying on a form. As needed, *mUzima* forms can incorporate data validation, branching logic, rule-based decision support, and multimedia items (Figure 4E). Data captured within forms can be saved as complete or as draft for future completion. A configurable autosave feature also exists to ensure that data collected in forms are not inadvertently lost.

Forms are automatically downloaded into the mobile application based on their inclusion within a program, but mobile device users can easily add or remove individual forms as per the user's preferences. With *mUzima*, there is no need to install a new version of the application every time new forms or form updates are available. Further, the *mUzima* form update mechanism is optimized to ensure that no data collected in older forms within the mobile application are lost during the form update process.

Registration and Patient Matching

Registration forms within *mUzima* rely on the form features. Prior to registering any new client, the application runs a probabilistic patient-matching algorithm to ensure that duplicate patients are not created. When no patient matches are found on the mobile device, *mUzima* runs a match against the server if there is connectivity. In settings where the device is offline and new individuals still need to be registered offline, the *mUzima* server-side module provides functionality to first match the new registration against existing EHR records once the registration information is uploaded to the server. If duplicates are found, functionality exists to merge the patient's records.

Geolocation Services

Patient tracing is a core component of community-based care, especially in areas without reliable address systems. Geolocation functionality within *mUzima* can help in finding patients within communities (Figure 4G). *mUzima* provides the ability to add and edit the GPS locations for individuals and employs navigation functionality to help locate patients. Functionality also exists to automatically capture the location where a clinical form is completed, should the program decide to enable this feature.

Relationship Support

The relationships feature captures the nature of relationships of various individuals with the reference client. Relationships are particularly relevant for contact tracing in the age of COVID-19 and for HIV index case testing services. In both cases, individuals related to or who have contact with the index case need to be traced for testing and follow-up. *mUzima* supports capturing information on the nature of the relationship between 2 individuals and collection of data on the related individual (Multimedia Appendix 3).

Table 1 summarizes key features within the *mUzima* application, highlighting often-overlooked elements in developing configurable and scalable mHealth applications for use in LMICs [36].

Table 1. *mUzima* features.

Feature	Description
EHR[a] system compatibility	*mUzima* was designed to be the mHealth[b] extension to a largely deployed EHR system (OpenMRS) to prevent siloed data collected within *mUzima*. OpenMRS is deployed nationally in several countries.
Security	Security features in *mUzima* include password-based log-in, data encryption, secure data transmission, timed user logouts, and a password-changing mechanism within limited connectivity settings.
Multiple use cases	*mUzima* is easily customizable to support any clinical use case, both within and outside of clinical facilities.
Data collection tools	*mUzima* uses easy-to-develop, web-based forms to collect data (including by providers off-site) that are securely transmitted and stored in an EHR system, as opposed to a siloed server
Offline capabilities	*mUzima* functions smoothly in both online and offline modes. This allows for data collection and review, even when the mobile device is offline.
Error resolution	*mUzima* prevents duplicate form data entry on mobile devices and resolves errors in the EHR system during data processing.
Form management	*mUzima* enables easy download and update of new forms onto the app without the need to re-install or restart the app. Form default text sizes are configurable, with the ability to magnify forms during use.
Cohort management	*mUzima* allows for the management of different patient or client groups on a mobile device.
Multiple languages	*mUzima* provides localization support for use in various languages. The full application is also currently available in 6 languages: English, Portuguese, French, Swahili, Gujarati, and Hindi.
Multiple themes	*mUzima* is shipped on 2 themes as of v2.5.0, including a dark theme (white on black) and light theme (black on white)
Relationships	In the health care space, a relationship is used to pair or associate 2 people whose care may be interlinked. *mUzima* supports recording of relationships (Multimedia Appendix 3).
Geomapping	*mUzima* can easily capture and record patient GPS coordinates and provides navigation services to support tracing of clients (Multimedia Appendix 4).
Clinical summary or abstracts	*mUzima* utilizes the OpenMRS Reporting Module to design and generate HTML reports that are then rendered on the app, providing comprehensive abstracts for each patient (Multimedia Appendix 4).

[a]EHR: electronic health record.

[b]mHealth: mobile health.

Items 6: Usability and Content Testing

In line with the Digital Development Principles of "Building for Sustainability" and "Be Collaborative," *mUzima* relies on a community of developers and implementers for content and user acceptance testing. As of February 2021, the *mUzima* community had 112 members, with a total of 31 unique contributors to the *mUzima* codebase. This community has made over 3800 contributions or commits to the codebase. Over 250 individuals have used, contributed, or reviewed *mUzima* documentation through *mUzima*'s Wiki platform [37], and more than 190 have contributed in the *mUzima* discussion forums [38]. The community is made up of stakeholders supported through various funded projects and agencies.

User testing for *mUzima* is conducted by a core team, as well as by community members prior to any *mUzima* release. Each feature in an *mUzima* release is documented as a JIRA ticket [39,40]. These tickets are used to track testing findings, with changes made based on feedback received. Final acceptance testing is conducted prior to deploying a new version of the application. Multimedia Appendix 5 provides a screenshot of release testing for a ticket in *mUzima* version 2.7.0, with details of testing status for each ticket in this version shared in a publicly available release plan [40].

Item 7: User Feedback

A user-centered design approach is employed in developing core *mUzima* functionality and features and in adapting the application for various use cases. An engaged Implementer Community provides suggestions for features, as well as feedback on developed application components and content. This feedback is primarily gathered via weekly community calls and through the *mUzima* forum [38]. Additional user inputs are gathered through reviews of the application using the feedback mechanism in Google Play Store, where *mUzima* has thousands of downloads and a 5-star rating [41].

Item 8: Access by Individual Participants

mUzima increases access to clinical data for providers who work outside of care facilities and in settings disconnected from the EHR system server. It is a provider-facing application and not available for direct use by patients. The application can leverage both WiFi-based or phone-based data bundles for data transmission. As an Android application, *mUzima* is not available to health providers who do not have Android smartphones and has limited utility to providers who are illiterate or unfamiliar with using smartphones. To use *mUzima*, providers need to have credentials to access an OpenMRS EHR instance, and this will limit utility of *mUzima* for those without such access or those using other EHR systems.

Item 9: Cost Assessment

A formal cost assessment has yet to be done on *mUzima*. Although *mUzima* is available free of charge to users, there are still several associated costs to implement it. Key costs include customization costs that involve creating relevant forms for a program; development of queries that define cohorts of patients; smartphone purchase; costs associated with hosting an OpenMRS instance (if one does not exist); connectivity costs to the server through WiFi or Internet; and costs of personnel to support users, maintain infrastructure, and conduct training. Comparatively, in instances where care is conducted outside of facilities, *mUzima* implementation should be cheaper than using laptops equipped with OpenMRS—given the higher cost associated with laptops and managing multiple OpenMRS instances. Although paper-based systems have historically been used to support care in LMICs, they have recurrent costs given that pieces of paper can only be used once to record data, and there are personnel costs associated with retrospective entry of data collected on paper into the EHR system [42]

Item 10: Adoption Inputs and Program Entry

mUzima is freely available on the Google Play Store [41], with all changes and enhancements in any new version shared broadly via multiple channels, as well as fully documented in the collaborative wiki space [37]. Within the wiki space are documents on setting up the system, as well as training materials. Multiple introductory and educational videos on *mUzima* are also available through the *mUzima* YouTube channel [43]. *mUzima* training often involves technical personnel as well as training of end user providers. Technical team members, comprised of developers and implementing information technology personnel, often need to understand how to customize *mUzima* and to set up the server and smartphone devices for use. The technical team training is largely hands-on, with the *mUzima* community forums and weekly meetings available for answering questions from community members. End users are often trained by the technical team members from their organization. These trainings typically cover core *mUzima* features and also the customized elements relevant to that program. For most end users, training on *mUzima* typically takes 1 to 2 days. Refresher trainings are available especially when there are upgrades to the application.

Item 11: Limitations for Delivery at Scale

The Digital Development Principle "Design for scale" is highly relevant to *mUzima*. Scalability was considered at the outset, allowing for the system to support any type of clinical condition, various workflows, and multiple versions of both OpenMRS and Android. Load testing is a core part of all *mUzima* version releases to ensure that *mUzima* works with larger patient populations and data needs. *mUzima's* ability to scale is evidenced by its successful large-scale adoption across multiple clinical settings and countries (Table 2). As a testament of its scalability and implementation success, *mUzima* has been recognized as a digital global public good since 2019 [22] and is listed as a global project in the WHO "Digital Health Atlas" [44].

Despite having scaled to national levels, further expansion of *mUzima* is limited by the fact that it only currently extends 1 EHR system, OpenMRS. *mUzima's* FHIR resources are under development to allow seamless health information exchange with other FHIR-compliant systems [34]. Customization of *mUzima* requires knowledge of form programming using HTML and basic knowledge on creating cohorts and queries within OpenMRS—both of which can limit scaling of *mUzima* implementations. Further, as a native Android application, *mUzima* will not work with smartphones that run on different operating systems.

Table 2. Large-scale *mUzima* implementations.

Type of implementation	Description
HIV care	*mUzima* is nationally endorsed and implemented by the Kenya Ministry of Health (MoH) to support HIV testing and screening (HTS) [45]. In Kenya, the application is in use at over 220 public facilities, with more than 1000 providers using it on any given workday. To date, over 500,000 HTS visits have been conducted in Kenya using *mUzima*. In addition, *mUzima* is in use in Zambezia province in Mozambique for defaulter and lost-to-follow-up tracing and for preventative visits (Multimedia Appendix 6). The application has also been adapted for national rollout in support of HIV patient tracing activities in Uganda.
Cancer care	Partners in Health Rwanda, working closely with the Rwanda MoH and the Clinton Health Access Initiative, are using *mUzima* for cervical and breast cancer screening and referral. Currently implemented in 3 districts, the *mUzima* application has already been used to support data collection during screening of more than 4000 patients for breast and cervical cancer at 39 facilities. The goal is to extend *mUzima* application use to 15 districts in Rwanda.
Chronic disease management (CDM)	*mUzima* is in use at more than 70 facilities in Kenya as part of a task-shifting program to support CDM, including hypertension and diabetes care. Unlike other implementations that use *mUzima* in communities, these CDM implementations use *mUzima* at dispensaries and within other facilities across 6 counties (Uasin Gishu, Bungoma, Trans-Nzoia, Nandi, Kisumu, and Busia). To date, over 350 providers have used *mUzima* to record CDM encounters in more than 100,000 visits. The CDM program incorporates robust decision support features based on alerts and reminders, as well as educational media with a personalized display based on patient-specific data (Figure 4E).
COVID-19	At the request of Kenya's MoH, *mUzima* was adapted to support COVID-19 care, including screening, contact tracing, testing, and symptom monitoring. The COVID-19 forms used are based on guidelines provided by the World Health Organization (WHO) and use standard terms based on the Columbia International eHealth Laboratory (CIEL) dictionary [46]. The *mUzima* COVID-19 application leverages relationship and geomapping features for contact tracing and is available for adaptation beyond Kenya (Multimedia Appendix 7). *mUzima* is also in use by the MoH in Rwanda for COVID-19 management.

Item 12: Contextual Adaptability

mUzima has been used across various health domains (Table 2), with most use in primary contexts that have limited internet connectivity, especially in Sub-Saharan Africa. *mUzima* provides localization support for use in various languages and is currently available in 6 languages, namely English, Portuguese, French, Swahili, Gujarati, and Hindi. The application comes with multiple configurable settings that are not hard coded into the application. Examples of configurable settings include whether to capture GPS location of activity, default font size, auto-synchronization feature, time setting for automatic time-outs, form auto-save time interval, among other features. Forms, cohorts, and clinical observations within the application can be customized without any need for reprogramming.

Item 13: Replicability

Content related to *mUzima* is organized in the application's website that highlights key application features and provides a demonstration instance for those interested in exploring the application [21]. This website provides links to all other *mUzima*-related resources and documentation. Detailed documentation on *mUzima* is available on the wiki page and includes user guides, technical documentation, and implementation guides, among others [37]. Multimedia content such as demonstration videos is also available on the *mUzima* YouTube channel [43]. *mUzima* code is shared via GitHub [47], and as of February 2021, there had been 650 pull requests and over 120 forks to the *mUzima* code.

Item 14: Data Security

mUzima is developed to enable the primary clinical implementing partners to remain full custodians of all data collected and transmitted to OpenMRS through *mUzima*. As such, identifiable patient information is only accessible to those given access through the EHR system. In line with the Principle for Digital Development to "Address privacy and security," the *mUzima* application has incorporated several security features. Access to the *mUzima* is done through a username and password validated initially with the EHR system. Users who forget their password have to be re-authenticated within the EHR system by an administrator. *mUzima* contains a feature for automatic time-outs where users are logged out after a program-defined period of inactivity. Bidirectional exchange of data between *mUzima* and the EHR system is secured through use of an https protocol. All *mUzima*-related data are stored in an encrypted format in the mobile device.

Item 15: Compliance With National Guidelines or Regulatory Statutes

Implementations using *mUzima* have full control of locations of servers where patient data from *mUzima* are stored. It is advocated that these implementations comply with data residency guidelines and laws for the country of operation. Organizations that opt to use *mUzima* are advised to contact relevant statutory bodies within countries that oversee mHealth applications. *mUzima* allows users to configure forms, patient identifiers, and concept dictionary terms to be used in each implementation. Compliance with national guidelines for these items is thus at the discretion of the implementers but is highly advocated. Forms and data elements can be easily developed to align with approved forms in countries and using any national dictionary should one be available. As an example, implementations in Kenya leverage the standards-based Columbia International eHealth Laboratory (CIEL) concept dictionary [46].

Item 16: Fidelity of the Intervention

Performance of the *mUzima* application is gathered using the Firebase Crashlystics plugin [48]. These performance metrics indicate that it takes <5 seconds for the application to launch and <1 second to upload each completed form payload to the OpenMRS EHR system from within *mUzima*. Over the most recent 3-month period (December 2020 to February 2021), the percentage of *mUzima* users without crashes in any of their usage sessions was 92.88%. *mUzima* also uses Firebase Crashlystics to monitor locations where *mUzima* has been used over a period of time, as well as high-level details of the users, broken down by user-provided gender and age (Multimedia Appendix 8). For individualized and detailed usage statistics, *mUzima* has also been programmed to collect paradata through logs of key components of the application while it is in use. Although performance and workflow analytics can be conducted on the logged paradata, there is a need to create a user-friendly and personalized dashboard highlighting a user's work performance on *mUzima* that can be availed to support decision making around work patterns.

Discussion

Although EHR and mHealth solutions have been widely embraced in LMICs, there is often lack of seamless data exchange between widely deployed EHR systems and these mHealth solutions. To our knowledge, this is one of the only descriptions of a robust extension of a widely used EHR system in LMICs. Most other widely deployed mHealth applications primarily save data to their own servers and have challenges integrating with existing EHR systems. It will be several years before all LMICs have effective interoperability and health information exchange infrastructure [49,50]. In the meantime, a paradigm that advocates for comprehensive mobile extensions to existing EHR systems will help to reduce the widening digital divide that adversely impacts resource-limited settings.

For mHealth applications to be effective and scalable within resource-limited settings, they have to build on lessons learned in implementation and development over the years [16,19,35]. Integration of mHealth applications with EHR systems uncovers new challenges that deserve attention. In particular, challenges emerge around best approaches to manage user credentials across 2 systems and in disconnected settings, appropriate delivery of subsets of EHR system data to mobile devices that have limited storage, management of ever-changing cohort sizes and patient-level data while ensuring data transmission costs are contained, and development of tethered mHealth solutions that can adapt to multiple use cases without the need to reprogram the application.

The *mUzima* team is currently broadening its support for implementations within various countries to ensure responsiveness to needs. As countries mature with regard to health information exchange, the *mUzima* team has started implementing FHIR, starting with patient, person, encounter, and observation FHIR [34]. The application user interface is undergoing redesign and revamping with input from users. The aim is primarily to enhance the user experience and to ensure that key features are more easily accessible to users through the interface. The *mUzima* team is in the process of adding secure messaging and communication features and incorporating a more robust teleconsultation feature, beyond the store-and-forward teleconsultation functionality already incorporated in the application. Finally, the application's modularization feature is being revamped to enable easier incorporation of additional plug-ins to *mUzima*. The *mUzima* data analytics team is also working on reusable paradata analytics and visualization to provide needed information on provider work performance and engagement with the application [51].

Evaluation is important for any health information system. Usability and feasibility of *mUzima* have previously been assessed in an implementation to support nurses in managing hypertensive patients [52]. Ongoing *mUzima* evaluations include (1) evaluation of patient-specific, phone-generated reminders through *mUzima* for hypertension care in Kenya; (2) evaluating use of *mUzima*-facilitated monitoring of health worker performance in Western Kenya using logged paradata; and (3) evaluation of the *mUzima* mobile application for patient tracing and preventive visits in Mozambique. Priorities for future evaluations should include comprehensive assessment of costs and benefits of using the *mUzima* application.

In conclusion, greater emphasis needs to be placed on mHealth applications that extend reach of EHR systems within resource-limited settings to reduce the digital divide that has emerged with use of standalone EHR systems or mHealth applications. *mUzima* demonstrates how this can be done at scale, with evident adoption across countries and for various types of care programs.

Acknowledgments

This work was made possible by the support of the American people through the United States Agency for International Development (USAID; grant number 7200AA18CA00019), Norwegian Agencies for Development Cooperation under the Norwegian Programme for Capacity Development in Higher Education and Research for Development (NORHED) program (Norad: Project QZA-0484), and Medtronics Foundation. The contents are solely the responsibility of the authors and do not necessarily represent the official views of USAID, NORHED, the United States Government, or Medtronics Foundation. The authors thank the *mUzima* developer and implementer community members and numerous other contributors to this project.

Authors' Contributions

MCW conceptualized the study, acquired funding, was the project administrator, supervised the study, wrote the original draft of the manuscript, and edited the manuscript. SS, BM, SM, and NR were the core *mUzima* development and implementation team and contributed the writing, review, and editing of the manuscript. PC provided critical input on the manuscript and contributed to its writing and revision. AY conducted *mUzima* project management, performed quality assurance, and contributed to the manuscript writing, review, and editing.

Conflicts of Interest

This paper describes an application that has the development team primarily as the leaders in its writing.

Multimedia Appendix 1
<italic>mUzima</italic> Server-Side Setup Configuration.
[PNG File , 679 KB - jmir_v23i12e26381_app1.png]

Multimedia Appendix 2
<italic>mUzima</italic> Error Resolution Mechanism.
[PNG File , 156 KB - jmir_v23i12e26381_app2.png]

Multimedia Appendix 3
<italic>mUzima</italic> Relationship Features.
[PNG File , 1879 KB - jmir_v23i12e26381_app3.png]

Multimedia Appendix 4
<italic>mUzima</italic> Geo-mapping feature & Clinical Summary.
[PNG File , 1564 KB - jmir_v23i12e26381_app4.png]

Multimedia Appendix 5
Screenshot of Release Testing for a Ticket in <italic>mUzima</italic> Version 2.7.0.
[PNG File , 252 KB - jmir_v23i12e26381_app5.png]

Multimedia Appendix 6
<italic>mUzima</italic> for Patient Tracing in Mozambique.
[MP4 File (MP4 Video), 9789 KB - jmir_v23i12e26381_app6.mp4]

Multimedia Appendix 7
<italic>mUzima</italic> for COVID-19.
[MP4 File (MP4 Video), 5321 KB - jmir_v23i12e26381_app7.mp4]

Multimedia Appendix 8
Locations of <italic>mUzima</italic> App Usage.
[PNG File , 226 KB - jmir_v23i12e26381_app8.png]

References

1. Williams F, Boren S. The role of the electronic medical record (EMR) in care delivery development in developing countries: a systematic review. Inform Prim Care 2008 Jul 01;16(2):139-145 [FREE Full text] [doi: 10.14236/jhi.v16i2.685] [Medline: 18713530]
2. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]
3. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. Bull World Health Org 2012 May 01;90(5):332-340. [doi: 10.2471/blt.11.099820]
4. Muinga N, Magare S, Monda J, Kamau O, Houston S, Fraser H, et al. Implementing an open source electronic health record system in Kenyan health care facilities: case study. JMIR Med Inform 2018 Apr 18;6(2):e22 [FREE Full text] [doi: 10.2196/medinform.8403] [Medline: 29669709]
5. Implementation of the UgandaEMR: Results of a Security Assessment. MEASURE Evaluation. 2020. URL: https://www.measureevaluation.org/resources/publications/tr-20-413.html [accessed 2021-11-19]
6. Chaplin B, Meloni S, Eisen G, Jolayemi T, Banigbe B, Adeola J, et al. Scale-up of networked HIV treatment in Nigeria: creation of an integrated electronic medical records system. Int J Med Inform 2015 Jan;84(1):58-68. [doi: 10.1016/j.ijmedinf.2014.09.006] [Medline: 25301692]
7. eSaúde: Building a Local OpenMRS Community to Support a National EMR Implementation in Mozambique. OpenMRS. 2016 Aug 31. URL: https://openmrs.org/2016/08/31/mozambique-case-study/ [accessed 2021-11-19]
8. Nsanzimana S, Kanters S, Remera E, Forrest JI, Binagwaho A, Condo J, et al. HIV care continuum in Rwanda: a cross-sectional analysis of the national programme. The Lancet HIV 2015 May;2(5):e208-e215. [doi: 10.1016/s2352-3018(15)00024-7]
9. Fast Track: ending the AIDS epidemic by 2030. UNAIDS. 2014. URL: https://www.unaids.org/sites/default/files/media_asset/JC2686_WAD2014report_en.pdf [accessed 2021-11-19]
10. Levi J, Raymond A, Pozniak A, Vernazza P, Kohler P, Hill A. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. BMJ Glob Health 2016 Sep 15;1(2):e000010 [FREE Full text] [doi: 10.1136/bmjgh-2015-000010] [Medline: 28588933]
11. Kang'a S, Puttkammer N, Wanyee S, Kimanga D, Madrano J, Muthee V, et al. A national standards-based assessment on functionality of electronic medical records systems used in Kenyan public-Sector health facilities. Int J Med Inform 2017 Jan;97:68-75. [doi: 10.1016/j.ijmedinf.2016.09.013] [Medline: 27919397]
12. Tweya H, Feldacker C, Gadabu OJ, Ng'ambi W, Mumba SL, Phiri D, et al. Developing a point-of-care electronic medical record system for TB/HIV co-infected patients: experiences from Lighthouse Trust, Lilongwe, Malawi. BMC Res Notes 2016 Mar 05;9(1):146 [FREE Full text] [doi: 10.1186/s13104-016-1943-4] [Medline: 26945749]
13. Were MC, Emenyonu N, Achieng M, Shen C, Ssali J, Masaba JPM, et al. Evaluating a scalable model for implementing electronic health records in resource-limited settings. J Am Med Inform Assoc 2010 May 01;17(3):237-244 [FREE Full text] [doi: 10.1136/jamia.2009.002303] [Medline: 20442140]
14. Muthee V, Bochner A, Kang'a S, Owiso G, Akhwale W, Wanyee S, et al. Site readiness assessment preceding the implementation of a HIV care and treatment electronic medical record system in Kenya. Int J Med Inform 2018 Jan;109:23-29. [doi: 10.1016/j.ijmedinf.2017.10.019] [Medline: 29195702]
15. Were M, Sinha C, Catalani C. A systematic approach to equity assessment for digital health interventions: case example of mobile personal health records. J Am Med Inform Assoc 2019 Aug 01;26(8-9):884-890 [FREE Full text] [doi: 10.1093/jamia/ocz071] [Medline: 31188438]

16. Lopéz DM, Blobel B. mHealth in low- and middle-income countries: status, requirements and strategies. Stud Health Technol Inform 2015;211:79-87. [Medline: 25980851]
17. Garner SL, Sudia T, Rachaprolu S. Smart phone accessibility and mHealth use in a limited resource setting. Int J Nurs Pract 2018 Feb 20;24(1):e12609. [doi: 10.1111/ijn.12609] [Medline: 29159919]
18. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
19. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
20. Health Information Systems Interoperability Maturity Toolkit. MEASURE Evaluation. URL: https://www.measureevaluation.org/resources/tools/health-information-systems-interoperability-toolkit [accessed 2021-11-19]
21. mUzima. URL: https://www.muzima.org/ [accessed 2021-11-19]
22. Digital Public Goods Definition. Digital Public Goods Alliance. URL: https://digitalpublicgoods.net/registry/ [accessed 2020-10-01]
23. 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. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
24. Principles for Digital Development. URL: https://digitalprinciples.org/ [accessed 2021-11-19]
25. OpenMRS. URL: https://openmrs.org/ [accessed 2021-11-19]
26. Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. AMIA Annu Symp Proc 2006:529-533 [FREE Full text] [Medline: 17238397]
27. OpenMRS Atlas. OpenMRS. URL: https://atlas.openmrs.org/ [accessed 2021-11-19]
28. Mobile Operating System Market Share Africa. Statcounter. URL: https://gs.statcounter.com/os-market-share/mobile/africa [accessed 2021-11-19]
29. Mozilla Public License Version 2.0. Mozilla Foundation. URL: https://www.mozilla.org/en-US/MPL/2.0/ [accessed 2021-11-19]
30. Apache Lucene. URL: https://lucene.apache.org [accessed 2020-10-01]
31. google / guice. GitHub. URL: https://github.com/google/guice [accessed 2020-10-01]
32. Indexer (programming). Wikipedia. URL: https://en.wikipedia.org/wiki/Indexer_(programming) [accessed 2021-11-19]
33. RESTful Services - HTTP. Sagar Mane. 2017. URL: https://medium.com/@sagar.mane006/restful-services-http-a6a83ca8fb7f [accessed 2020-10-01]
34. Mandel J, 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;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
35. WHO Guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/reproductivehealth/publications/digital-interventions-health-system-strengthening/en/ [accessed 2021-11-19]
36. Features. mUzima. URL: https://www.muzima.org/features/ [accessed 2020-10-01]
37. muzima. mUzima Wiki. URL: https://wiki.muzima.org/ [accessed 2020-10-01]
38. mUzima Forums. mUzima. URL: https://forums.muzima.org [accessed 2020-10-01]
39. JIRA Software. Atlassian. URL: https://www.atlassian.com/software/jira [accessed 2021-11-19]
40. mUzima 2.7.0 Release plan. mUzima Wiki. 2020 Dec 08. URL: https://wiki.muzima.org/display/muzima/mUzima+2.7.0+Release+plan [accessed 2021-11-19]
41. mUzima. Google Play Store. URL: https://play.google.com/store/apps/details?id=com.muzima&hl=en&gl=US [accessed 2020-10-01]
42. Were M, Kariuki J, Chepng'eno V, Wandabwa M, Ndege S, Braitstein P, et al. Leapfrogging paper-based records using handheld technology: experience from Western Kenya. Stud Health Technol Inform 2010;160(Pt 1):525-529. [Medline: 20841742]
43. Introducing mUzima for COVID-19. youtube. 2020 Jul 12. URL: https://www.youtube.com/watch?v=yDQRijQjIXA&t=2s [accessed 2020-10-01]
44. mUzima. Digital Health Atlas. URL: https://digitalhealthatlas.org/en/-/projects/1437/published [accessed 2021-11-19]
45. Farmer B. What The U.S. Could Learn From Vanderbilt's Electronic Health Records Experiment In Kenya. WPLN News. 2018 Jan 29. URL: https://wpln.org/post/what-the-u-s-could-learn-from-vanderbilts-electronic-health-records-experiment-in-kenya/ [accessed 2021-11-19]
46. Blaya J, Mamlin B. The MVP-CIEL Concept Dictionary. OpenMRS Wiki. 2020 Oct 5. URL: https://wiki.openmrs.org/display/docs/The+MVP-CIEL+concept+dictionary [accessed 2021-11-19]
47. mUzima. GitHub. URL: https://github.com/muzima [accessed 2020-10-01]

48. Firebase Crashlytics. Firebase. URL: https://firebase.google.com/docs/crashlytics [accessed 2020-10-01]
49. Kumar M, Mostafa J. Research evidence on strategies enabling integration of electronic health records in the health care systems of low- and middle-income countries: A literature review. Int J Health Plann Manage 2019 Apr 14;34(2):e1016-e1025. [doi: 10.1002/hpm.2754] [Medline: 30762907]
50. Higman S, Dwivedi V, Nsaghurwe A, Busiga M, Sotter Rulagirwa H, Smith D, et al. Designing interoperable health information systems using Enterprise Architecture approach in resource-limited countries: A literature review. Int J Health Plann Manage 2019 Jan 05;34(1):e85-e99. [doi: 10.1002/hpm.2634] [Medline: 30182517]
51. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. Mhealth 2020 Jan;6:7-7 [FREE Full text] [doi: 10.21037/mhealth.2019.10.01] [Medline: 32190618]
52. Vedanthan R, Blank E, Tuikong N, Kamano J, Misoi L, Tulienge D, et al. Usability and feasibility of a tablet-based Decision-Support and Integrated Record-keeping (DESIRE) tool in the nurse management of hypertension in rural western Kenya. Int J Med Inform 2015 Mar;84(3):207-219 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.12.005] [Medline: 25612791]

Abbreviations

API: application programming interface
APK: Android application package
CIEL: The Columbia International eHealth Laboratory
EHR: electronic health record systems
FHIR: fast health care interoperability resources
HIS: health information system
LMICs: low- and middle-income countries
mERA: mHealth Evidence Reporting and Assessment
mHealth: mobile health
MoH: Ministry of Health
NORHED: Norwegian Programme for Capacity Development in Higher Education and Research for Development
REST: representational state transfer
USAID: United States Agency for International Development
WHO: World Health Organization

Edited by R Kukafka, G Eysenbach; submitted 09.12.20; peer-reviewed by N Puttkammer, J Florez-Arango, C Paton; comments to author 08.02.21; revised version received 22.02.21; accepted 09.10.21; published 14.12.21.

Please cite as:
Were MC, Savai S, Mokaya B, Mbugua S, Ribeka N, Cholli P, Yeung A
mUzima Mobile Electronic Health Record (EHR) System: Development and Implementation at Scale
J Med Internet Res 2021;23(12):e26381
URL: https://www.jmir.org/2021/12/e26381
doi:10.2196/26381
PMID:34904952

Original Paper

Short Form of the Pediatric Symptom Checklist-Youth Self-Report (PSC-17-Y): Spanish Validation Study

Jose A Piqueras[1*], PhD; Verónica Vidal-Arenas[2*], MS; Raquel Falcó[1*], MS; Beatriz Moreno-Amador[1*], MS; Juan C Marzo[3*], PhD; Juliana M Holcomb[4*], BA; Michael Murphy[5*], EdD

[1]Area of Personality, Assessment and Psychological Treatment, Department of Health Psychology, Universidad Miguel Hernandez de Elche, Elche, Spain

[2]Department of Basic and Clinical Psychology and Psychobiology, Universitat Jaume I, Castellon, Spain

[3]Area of Social Psychology, Department of Health Psychology, Universidad Miguel Hernandez de Elche, Elche, Spain

[4]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[5]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Jose A Piqueras, PhD
Area of Personality, Assessment and Psychological Treatment
Department of Health Psychology
Universidad Miguel Hernandez de Elche
Edificio Altamira
Avenida de la Universidad, s/n
Elche, 03202
Spain
Phone: 34 966658343
Fax: 34 966658904
Email: jpiqueras@umh.es

Abstract

Background: The short form, 17-item version of the Pediatric Symptom Checklist-Youth Self-Report (PSC-17-Y) is a validated measure that assesses psychosocial problems overall (OVR) and in 3 major psychopathological domains (internalizing, externalizing, and attention-deficit/hyperactivity disorder), taking 5-10 min to complete. Prior research has established sound psychometric properties of the PSC-17-Y for English speakers.

Objective: This study extends psychometric evidence for the acceptability of the PSC-17-Y in a large sample of Spanish adolescents, providing proof of its reliability and structure, convergent and discriminant validity, and longitudinal and gender invariance.

Methods: Data were collected on 5430 adolescents, aged 12-18 years, who filled out the PSC-17-Y twice during 2018-2019 (7-month interval). We calculated the Cronbach alpha and the McDonald omega coefficients to test reliability, the Pearson correlation for convergent (distress) and criterion validity (well-being, quality of life, and socioemotional skills), confirmatory factor analysis (CFA) for structure validity, and multigroup and longitudinal measurement invariance analysis for longitudinal and gender stability.

Results: Within structural analysis for the PSC-17-Y, CFA supported a correlated 3-factor solution, which was also invariant longitudinally and across gender. All 3 subscales showed evidence of reliability, with coefficients near or above .70. Moreover, scores of PSC-17-Y subscales were positively related with convergent measures and negatively related with criterion measures. Normative data for the PSC-17-Y are presented in the form of percentiles (75th and 90th).

Conclusions: This work provides the first evidence of the reliability and validity of the Spanish version of the PSC-17-Y administered over the internet to assess mental health problems among adolescents, maintaining the same domains as the long version.

(J Med Internet Res 2021;23(12):e31127) doi:10.2196/31127

KEYWORDS

PSC-17-Y; psychometric properties; screening; mental problems; adolescents; adolescent health; adolescent medicine; psychiatry; psychology; psychosocial issues

Introduction

According to Polanczyk et al [1], the most common mental health disorders among children and adolescents include anxiety or depression, behavioral disorders, and attention-deficit/hyperactivity disorder (ADHD). Emotional and behavioral symptoms at the subclinical level raise the risk of subsequent development of mental disorders [2]. Moreover, the COVID-19 pandemic has provoked a considerable increase in mental health problems among children and adolescents [3-5].

National and international policies and strategies globally recommend that young people attending primary care should be routinely screened for psychosocial problems [6]. Despite this, such screening occurs in less than 50% of primary care visits of adolescents, meaning that more than half of adolescent mental health problems go undetected [7,8]. Although several screening tools exist for psychosocial problems in young people, most cover a single domain [9] and can be time consuming to administer and interpret [10]. Primary care clinicians can often be unsure of which screening tools are appropriate for their clinical context. In addition, many tools rely on the provider having the skills, knowledge, expertise, and experience to initiate screening, interpret results, and provide appropriate interventions [8]. Providers often describe a lack of resources in terms of the availability of time, appropriate tools, training, and experience in youth health [11].

A recent review of "Self-Report Rating Scales to Guide Measurement-Based Care in Child and Adolescent Psychiatry" [12] highlights that the Pediatric Symptom Checklist (PSC) is 1 of the most widely used measures to screen psychosocial problems in primary care units and school settings. This statement is supported for all parent and youth reports and for long and short forms (parent- and youth-reported long form [PSC-35]; parent- and youth-reported short form [PSC-17]) [13-23].

The short form, 17-item version of the PSC-Y (PSC-17-Y) [16] is used to assess self-reported general psychosocial functioning among youth above 11 years old, taking only 5-10 min to be completed, and is statistically equivalent to the short form of the parent version (PSC-17) [16] and to the longer youth report form (PSC-35-Y [19-21]).

Three studies of the parent report PSC-17 (Gardner et al [16,17] and Murphy et al [18]) have confirmed the existence of the 3 original subscales for internalizing (INT) symptoms, externalizing (EXT) symptoms, and ADHD symptoms (ATT) and provided evidence of the reliability of the overall (OVR) scale. Two studies with the youth-reported short form have been published. On the one hand, Bergman et al [22] found that the PSC-17-Y is equivalent to the parent-reported form of the PSC-17, indicating that a 3-factor short form with 17 items meets the criteria for scalar invariance across gender. On the other hand, Parker et al [23] examined the screening validity of the PSC-17-Y in a child welfare population. Youth with any lifetime mental health diagnosis scored significantly higher on the PSC ATT and INT subscales. The ATT, INT, and OVR subscale scores were significantly correlated with psychosis, depression, and anxiety disorder scores. ADHD was associated with ATT, OVR, and EXT scores. Only bipolar disorder was weakly associated with PSC subscale scores (EXT and OVR). This study provides support for the convergent and discriminant validity of the PSC-17-Y.

Despite PSC-17-Y's potential, however, there is limited evidence of some of its relevant psychometric properties (eg, longitudinal measurement invariance and other reliability coefficients different from the Cronbach alpha [α]), and to the best of our knowledge from the scientific literature review, none of these psychometric analyses are in languages other than English (ie, none on Spanish populations).

Thus, this work aimed to extend the psychometric evidence for the acceptability of the PSC-17-Y in a large sample of Spanish adolescents, providing different sources of reliability and validity. This research could facilitate the use of the PSC -17-Y in more contexts and for more possible applications in youth mental health settings. Overall, we expected that the PSC-17-Y would show that it is a valid and reliable ultrabrief screening measure that can be administered over the internet to detect mental health problems in Spanish adolescents.

Methods

Sample

The final sample consisted of 5430 adolescents (2769 [51%] females) at time 0 and 2117 (1109 [52.4%] females) at time 1 (approximately 7 months later). The participants were enrolled in Spanish secondary education grades, equivalent to US middle and high school, from grades 7 (12-13 years) to 12 (17-18 years). The average age of the sample at time 0 was 14.17 years (SD 1.50) and of the sample that participated at time 0 and time 1 was 13.99 years (SD 1.39).

Measures

PSC-17-Y

The PSC-17-Y [16] consists of 17 items and 3 factors to assess 3 types of problems: INT symptoms (ie, depression and anxiety), EXT symptoms (ie, disruptive behavior), and ATT, as well as an OVR score. The Spanish version of the PSC-17-Y was developed in accordance with the guidelines of the International Test Commission [24], using an iterative translation method that began with several independent translations. The item translations were then reviewed by a joint committee comprising translators with knowledge of the Spanish language and culture and specialists in the field of assessment who analyzed the adequacy of the adapted version. To be sure that adolescents properly understood all items, interviews asking about the comprehension of the items were conducted. In addition, in 2018, we had conversations with colleagues who had worked

on the translation of the PSC for parents in Chile [25,26] in order to obtain an adequate cross-cultural adaptation into the European Spanish language of the PSC-17-Y.

Social-Emotional Distress Survey-Secondary

The Social-Emotional Distress Survey-Secondary (SEDS-S) [27] is a 10-item behavioral screening questionnaire designed to measure INT distress. The reliability of the 1-factor total scale was α=.91. In their study, Dowdy et al [27] found a significant positive association of the SEDS-S distress factor with symptoms of anxiety and depression and a significant negative association with life satisfaction and strength scores.

Mental Health Continuum-Short Form

The Mental Health Continuum-Short Form (MHC-SF) [28] is the reduced version of the MHC Long Form. This measure provides self-reported well-being, divided into 3 subfactors: psychological (6 items), emotional (3 items) and social well-being (5 items). In this study, we used the Spanish version of the MHC-SF recently adapted by our team. The MHC-SF has received psychometric support for use with adolescents across many different countries, including Spain [29,30], showing excellent internal consistency (Cronbach α>.80) and discriminant validity in adolescents.

KIDSCREEN-10 Index

The KIDSCREEN-10 Index [30] is a unidimensional scale that measures health-related quality of life (HRQoL) in healthy and chronically ill children and adolescents. It was developed to specifically identify children at risk in terms of subjective health and suggest appropriate early interventions. The instrument provides an overall HRQoL index covering the physical, psychological, and social facets of the HRQoL. Internal consistency values (Cronbach α) reach .82, and test-retest reliability within 2 weeks reaches .55 [31].

Social-Emotional Health Survey-Secondary

The Social-Emotional Health Survey-Secondary (SEHS-S) [32] was developed to measure the components of the covitality latent construct among youth. We used the Spanish version of the SEHS-S, which is appropriate for adolescents aged 12-18 years [33]. The SEHS-S includes 36 items for the assessment of core psychosocial assets based on a higher-order model comprising 12 first-order, grouped into 4 second-order, latent traits (3 each) and a higher-order general factor (covitality). The first domain, called belief-in-self, measures self-efficacy, self-awareness, and persistence. The domain belief-in-others comprises school support, peer support, and family support. The domain emotional competence considers emotion regulation, empathy, and behavioral self-control. Engaged living, which is the final domain, comprises 3 subscales: gratitude, zest, and optimism.

Procedure

This research used a non-experimental, transversal/longitudinal, quantitative, and descriptive-correlational design [34,35]. The UMH Project Evaluation Committee approved the study (reference no. DPS.JPR.02.17). Once the project was approved, quota sampling was carried out in 2 areas of southeastern Spain: the province of Alicante (PA) belonging to the Valencian Community, and the Autonomous Community of the Region of Murcia (RM), making a random selection of secondary schools based on ownership (public/nonpublic schools) and regional geographical areas (9 areas in PA and 21 in RM). After 100 schools were contacted, 13 from PA and 21 from RM agreed to participate, resulting in a total of 34 secondary schools (22 [65.2%] public and 12 [34.8%] nonpublic schools; 30 [87%] secular and 4 [13%] catholic schools).

Once the schools agreed to participate, signed informed consent was requested in writing from the parents/legal guardians of the adolescent participants and from the adolescents themselves, accepting participation in the research. The data collection was carried out in the schools and supervised by the research staff in person. The self-reporting assessment protocol was individually applied through the online survey tool LimeSurvey (LimeSurvey GmbH, Hamburg, Germany). Participation was voluntary, and the adolescents did not receive any incentive for their collaboration, while each school received a feedback report, including results by class group.

Data Analysis

All analyses were conducted using IBM SPSS Statistics version 25 and Mplus 8.4 (Muthén & Muthén). Confirmatory factor analysis (CFA) was conducted to test the structural validity. Figure 1 represents the correlated 3-factor solution tested. We used a diagonally weighted least squares means and variance adjusted (WLSMV) model estimator due to a number of alternative responses and the nonnormality distribution [36]. We tested the model's goodness of fit using the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root-mean-square error of approximation (RMSEA). A CFI of >0.90 and a TLI of >0.95 indicate an acceptable and an optimal fit, respectively [37], and RMSEA values of ≤0.10 indicate an acceptable fit [38].

Later, we tested whether the PSC-17-Y exhibits an invariant structure across gender and across time through longitudinal and multigroup measurement invariance analysis. In particular, 3 levels of invariance were tested: (1) *configural* (test whether all items load on the proposed factor), (2) *metric* (test whether item-factor loadings are similar across groups), and (3) *scalar* (test whether unstandardized item thresholds are similar across groups). In addition, within longitudinal measurement, invariance residuals covariances between the same item over time (eg, time 0 with time 1) were included. Thus, to indicate a significant decrement in fit when testing for measurement invariance, we used model comparison criteria of ΔCFI/ΔTFI ≥ 0.01 (ie, a decrease indicating the worst fit) [39] and ΔRMSEA ≥ 0.015 (ie, an increase indicating the worst fit) [40].

When there is scalar measurement invariance, the comparison of factor means across groups is permissible [41]. Consequently, we calculated gender differences. We also estimated the Cohen d index (standardized mean difference), which allows evaluating the effect size (ES) of the obtained differences [42].

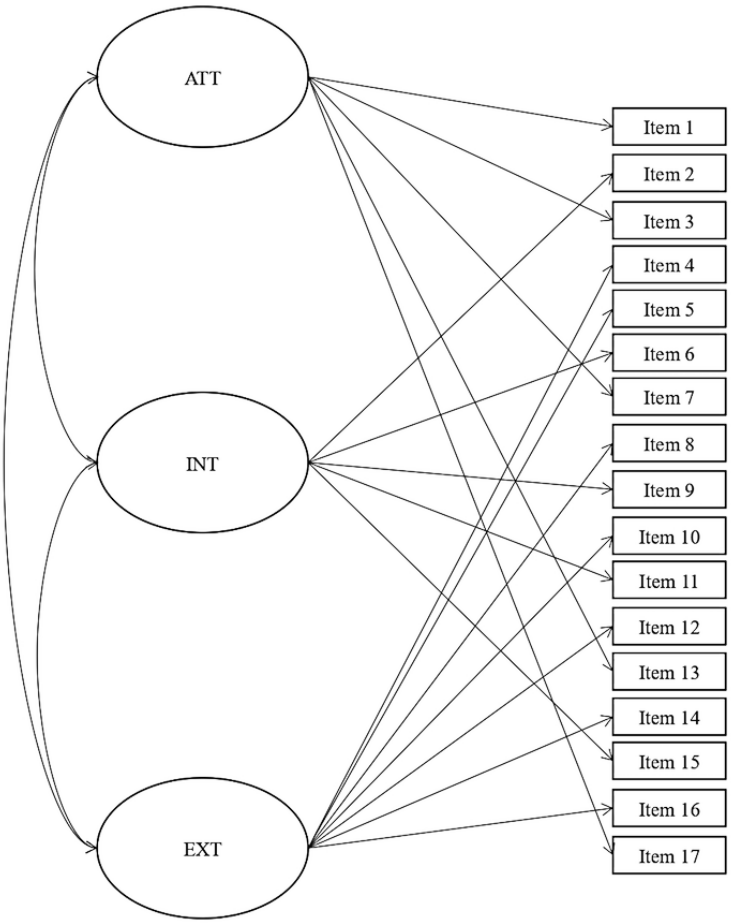
The Cronbach α and the McDonald omega (ω) [43] were used to estimate the internal consistency of the PSC-17-Y since the McDonald ω is a better estimator of reliability than the Cronbach α [44].

Convergent and criterion validity was evaluated by calculating the correlation coefficients between the scores on the PSC-17-Y and different, well-established measures. Specifically, we tested the convergent validity with measures of distress (SEDS-S) and criterion validity with measures of well-being, QoL, and socioemotional skills (MHC-SF, KIDSCREEN-10 Index, SEHS-S). The Cohen criteria were used to estimate the ES of the correlations [42,45].

Finally, normative data for the PSC-17-Y were presented in the form of percentiles (75th and 90th). We also calculated the cut-off point of 15 for OVR, 5 for INT symptoms, 7 for EXT symptoms, and 7 for ATT, as proposed by Gardner et al [16,17], because these scores have not received evidence-based support in Spanish adolescents.

As the sample size determination for psychometric validation studies lacks clear recommendations [46], we determined the required sample size by allocating several observations 5-10 times greater than the variables [47]. Accordingly, the sample size needed ranged between 85 and 170 participants based on the number of items in the PSC-17-Y. Furthermore, according to the subject-to-item ratio method, a sample size of ≥1000 to perform exploratory factor analysis (EFA) or CFA would be excellent [46].

Figure 1. Visual representation of the 17-item, 3-factor solution of the PSC-17-Y. ATT: attention-deficit/hyperactivity disorder symptoms; EXT: externalizing; INT: internalizing; PSC-17-Y: Pediatric Symptom Checklist-Youth Self-Report.



Results

Confirmatory Factor Analysis and Measurement Invariance

Findings of CFA and measurement invariance over time and across gender groups are summarized in Table 1. Since item 13 (“Me cuesta mucho cansarme”/“Acts as if driven by a motor”) had a low loading at the ATT subscale (0.127), we also tested the 16-item model, which slightly improved the fit. An optimal fit index was observed for both the 17- and 16-item models, with loadings from 0.348 (item 5) to 0.858 (item 6) and from 0.346 (item 5) to 0.859 (item 6), respectively. Longitudinal and multigroup measurement invariance was found for the 17- and 16-item models, which means that the structure, loadings, and intercepts of the PSC-17-Y were invariant over time and across gender groups. Although the 16-item version was slightly stronger than the 17-item version, we concluded that the minimal improvement in accuracy was less important than being able to use the 17-item model, which is the internationally recognized version and facilitates comparisons with other studies. For this reason, subsequent analyses were performed with a 17-item version.

Table 1. Goodness of fit for baseline models and measurement invariance of the 3-factor solution.

Number	Items, n	Overall goodness of fit				Comparative goodness of fit			
		χ^{2a} (*df*)	CFI[b]	TLI[c]	RMSEA[d] (90% CI)	Model comparison	ΔCFI	ΔTLI	ΔRMSEA
Baseline model									
1	17	1462.63 (116)	0.961	0.954	0.046 (0.044-0.048)	__[e]	—	—	—
2	16	1105.35 (101)	0.970	0.965	0.043 (0.041-0.045)	1 vs 2	0.009	0.010	–0.003
Longitudinal measurement configural invariance									
3	17	2814.41 (496)	0.956	0.950	0.029 (0 .028-0.030)	—	—	—	—
4	16	1957.05 (433)	0.970	0.966	0.025 (0 .024-0.027)	—	—	—	—
Longitudinal measurement metric invariance									
5	17	2395.55 (505)	0.964	0.960	0.026 (0 .025-0.027)	3 vs 5	0.008	0.010	–0.003
6	16	1845.45 (443)	0.973	0.969	0.024 (0 .023-0.025)	4 vs 6	0.003	0.003	–0.001
Longitudinal measurement scalar invariance									
7	17	2538.15 (522)	0.961	0.958	0.027 (0 .026-0.028)	5 vs 7	–0.003	–0.002	0.001
8	16	1969.06 (459)	0.970	0.968	0.025 (0 .024-0.026)	6 vs 8	–0.003	–0.001	0.001
Gender configural invariance									
9	17	2573.15 (990)	0.968	0.964	0.024 (0 .023-0.025)	—	—	—	—
10	16	2110.28 (866)	0.974	0.971	0.023 (0 .022-0.024)	—	—	—	—
Gender metric invariance									
11	17	2777.51 (1018)	0.965	0.961	0.025 (0 .024-0.026)	9 vs 11	–0.003	–0.003	0.001
12	16	2342.64 (892)	0.970	0.967	0.024 (0 .023-0.026)	10 vs 12	–0.004	–0.004	–0.001
Gender scalar invariance									
13	17	3068.45 (1046)	0.959	0.956	0.027 (0 .026-0.028)	11 vs 13	–0.006	–0.005	0.002
14	16	2460.40 (918)	0.968	0.966	0.025 (0 .024-0.026)	12 vs 14	–0.002	–0.001	0.001

[a]*P*<.001.

[b]CFI: comparative fit index.

[c]TLI: Tucker-Lewis index.

[d]RMSEA: root-mean-square error of approximation.

[e]Not applicable.

The correlation matrix of the items can be seen in Table 2.

As can be seen in Tables 3 and 4, the reliability coefficients for all measures ranged from .64 to .76 (McDonald ω). Regarding gender differences among total scores, females showed significantly higher scores on general social-emotional distress (d=0.34) and INT (d=0.42) scales than males, with a small-to-medium ES. Concerning the differences in the EXT scale, males showed more symptoms of EXT problems than females, with a small ES (d=0.12). Although the scores for well-being and socioemotional competencies were lower in females than in males, the ESs were low (d ranged from 0.08 to 0.15). Finally, males showed higher HRQoL levels than females, with an ES of 0.35.

According to Table 5, the intercorrelation between the PSC-17-Y subscales was moderate, indicating that this measure is composed of 3 differentiated and mutually associated factors. Regarding the relationship between PSC-17-Y subscales and the remaining measures, the highest positive correlations were between the INT subscale and the measure of distress (large ES), while the association of the ATT and EXT subscales with distress was medium. Similarly, the correlation was higher (and negative) between the INT subscale and the HRQoL (large ES) than between different types of well-being and socioemotional competencies (moderate to large). However, the correlation between ATT and measures of well-being, socioemotional competencies, and the HRQoL was significant and negative, with a small-to-medium magnitude. Finally, the EXT subscale presented a small-to-medium association with the different positive measures.

Table 2. Correlation matrix of the items.

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	__[a]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	0.15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
3	0.28	0.33	—	—	—	—	—	—	—	—	—	—	—	—	—	—
4	0.14	0.16	0.19	—	—	—	—	—	—	—	—	—	—	—	—	—
5	0.15	0.10	0.15	0.32	—	—	—	—	—	—	—	—	—	—	—	—
6	0.10	0.73	0.33	0.25	0.17	—	—	—	—	—	—	—	—	—	—	—
7	0.29	0.38	0.49	0.20	0.20	0.41	—	—	—	—	—	—	—	—	—	—
8	0.21	0.39	0.28	0.24	0.16	0.36	0.32	—	—	—	—	—	—	—	—	—
9	0.13	0.52	0.27	0.16	0.08	0.56	0.32	0.33	—	—	—	—	—	—	—	—
10	0.18	0.30	0.25	0.36	0.22	0.32	0.30	0.35	0.31	—	—	—	—	—	—	—
11	0.13	0.58	0.25	0.18	0.155	0.60	0.35	0.28	0.45	0.28	—	—	—	—	—	—
12	0.31	0.30	0.40	0.29	0.227	0.32	0.41	0.44	0.23	0.38	0.27	—	—	—	—	—
13	0.22	–0.04	0.05	0.07	0.105	0.06	0.06	0.06	–0.01	0.05	–0.01	0.16	—	—	—	—
14	0.26	0.24	0.34	0.31	0.256	0.24	0.33	0.41	0.20	0.41	0.19	0.48	0.14	—	—	—
15	0.13	0.42	0.20	0.07	0.026	0.41	0.30	0.23	0.47	0.18	0.36	0.17	0.07	0.14	—	—
16	.15	.24	.24	.29	.174	.28	.25	.37	.13	.34	.16	.43	.11	.42	.07	—
17	0.33	0.31	0.576	0.178	0.187	0.34	0.69	0.33	0.27	0.28	0.28	0.44	0.07	0.37	0.26	0.30

[a]Not applicable.

Table 3. Descriptive analysis for males and females and reliability coefficients of the PSC-17-Y[a].

PSC-17-Y Subscales	Cronbach α (95% CI)		Interitem correlations			McDonald ω (95% CI)		Score, mean (SD)		Effect size (d)
	Female	Male	Mean	Min	Max	Female	Male	Female (x)	Male (y)	d = x – y (*P* value)
ATT[b]	.61 (0.59-0.63)	.63 (0.61-0.65)	0.306	0.050	0.686	.65 (0.63-0.67)	.64 (0.62-0.66)	4.45 (1.98)	4.46 (2.07)	0.01 (.76)
INT[c] symptoms	.76 (0.75-0.78)	.73 (0.71-0.74)	0.510	0.412	0.730	.76 (0.75-0.78)	.72 (0.70-0.74)	3.74 (2.37)	2.79 (2.17)	0.42 (<.001)
EXT[d] symptoms	.65 (0.63-0.67)	.67 (0.65-0.68)	0.327	0.161	0.348	.66 (0.64-0.68)	.67 (0.65-0.69)	2.54 (2.07)	2.80 (2.15)	–0.12 (<.001)

[a]PSC-17-Y: short form of the Pediatric Symptom Checklist-Youth Self-Report.

[b]ATT: attention-deficit/hyperactivity disorder symptoms.

[c]INT: internalizing.

[d]EXT: externalizing.

Table 4. Descriptive analysis for males and females and reliability coefficients of study measures.

Measures	Cronbach α (95% CI)		McDonald ω (95% CI)		Score, mean (SD)		Effect size (d)
	Female	Male	Female	Male	Female (x)	Male (y)	d = x – y (*P* value)
Emotional	.80 (0.79-0.81)	.77 (0.75-0.78)	.82 (0.80- 0.83)	.79 (0.77-0.81)	13.84 (3.29)	14.31 (3.06)	–0.15 (<.001)
Social	.85 (0.84-0.86)	.83 (0.82-0.84)	.86 (0.85-0.86)	.83 (0.82-0.84)	19.67 (5.60)	20.36 (5.50)	–0.12 (<.001)
Psychological	.86 (0.86-0.87)	.86 (0.85-0.87)	.91 (0.85-0.87)	.86 (0.85-0.87)	28.04 (5.74)	28.51 (5.58)	–0.08 (.003)
HRQoL[a]	.86 (0.85-0.86)	.83 (0.82-0.84)	.86 (0.85-0.86)	.83 (0.82-0.84)	37.39 (7.36)	39.81 (6.44)	–0.35 (<.001)
Socioemotional skills	.91 (0.91-0.92)	.90 (0.90-0.91)	.91 (0.91-0.92)	.90 (0.90-0.91)	110.69 (14.72)	112.28 (14.04)	–0.12 (<.001)

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

Table 5. Correlations between the PSC-17-Y[a], distress, and well-being measures.

PSC-17-Y Subscales – Measures	ATT[b]	INT[c] symptoms	EXT[d] symptoms	Distress	Emotional well-being	Social well-being	Psychological well-being	HRQoL[e]	Socioemotional skills
ATT	1	—[f]	—	—	—	—	—	—	—
INT symptoms	0.35[g]	1	—	—	—	—	—	—	—
EXT symptoms	0.46[g]	0.36[g]	1	—	—	—	—	—	—
Distress	0.35[g]	0.67[g]	0.35[g]	1	—	—	—	—	—
Emotional well-being	–0.19[g]	–0.51[g]	–0.24[g]	–0.44[g]	1	—	—	—	—
Social well-being	–0.21[g]	–0.46[g]	–0.26[g]	–0.38[g]	0.65[g]	1	—	—	—
Psychological well-being	–0.23[g]	–0.50[g]	–0.31[g]	–0.42[g]	0.70[g]	0.72[g]	1	—	—
HRQoL	–0.30[g]	–0.67[g]	–0.34[g]	–0.60[g]	0.66[g]	0.63[g]	0.72[g]	1	—
Socioemotional skills	–0.31[g]	–0.45[g]	–0.41[g]	–0.39[g]	0.60[g]	0.61[g]	0.70[g]	0.67[g]	1

[a]PSC-17-Y: short form of the Pediatric Symptom Checklist-Youth Self-Report.

[b]ATT: attention-deficit/hyperactivity disorder symptoms.

[c]INT: internalizing.

[d]EXT: externalizing.

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

[f]Not applicable.

[g]$P<.001$.

Concerning normative information for PSC-17-Y subscales, each subscale is scored by the sum of its items. The adolescent's score on the scale can then be used to obtain the corresponding percentile score. The normative information for each of the 3 PSC-17-Y subscales and total scores for the whole sample are shown in Table 6. We also included percentile scores for PSC-17-Y OVR scores to facilitate international comparisons.

The values obtained in our study using the international cut-off point of 15 for OVR, 7 for ATT, 5 for INT symptoms, and 7 for EXT symptoms, as proposed by Gardner et al [24], were 20.7%, 15.1%, 29.7%, and 5.1%, respectively. The 90th percentile indicated that 11.6% of the sample scored above this cut-off point of 17 for OVR on the PSC-17-Y, 26.3% of the participants exceeded the 75th percentile for the PSC-17-Y total score. Regarding specific symptoms, 15.1%, 10.2%, and 10.6% of the participants scored above the 90th percentile on ATT, INT symptoms, and EXT symptoms, respectively. The data at the 75th percentile or quartile 1 indicated that 29.1%, 28.7%, and 32.0% of the participants exceeded the cut-off points for ATT, INT symptoms, and EXT symptoms, respectively.

Table 6. Normative information about PSC-17-Y[a] scales for adolescents (percentile scores); N=5430 (boys and girls 12-18 years old).

Percentile	PSC-17-Y ATT[b]	PSC-17-Y INT[c] symptoms	PSC-17-Y EXT[d] symptoms	PSC-17-Y total[e]
1	0	0	0	0
5	1	0	0	3
10	2	1	0	4
15	2	1	0	5
20	3	1	1	6
25	3	1	1	7
30	3	2	1	8
35	4	2	2	8
40	4	2	2	9
45	4	3	2	10
50	5	3	2	10
55	5	3	3	11
60	5	4	3	12
65	5	4	3	12
70	5	4	4	13
75	6	5	4	14
80	6	5	4	15
85	7	6	5	16
90	7	7	6	17
95	8	8	7	19
99	9	10	8	23

[a]PSC-17-Y: short form of the Pediatric Symptom Checklist-Youth Self-Report.

[b]ATT: attention-deficit/hyperactivity disorder symptoms. Rating anchor: ATT=0-10 (5 items).

[c]INT: internalizing. Rating anchor: INT symptoms=0-10 (5 items).

[d]EXT: externalizing. Rating anchor: EXT symptoms=0-14 (7 items).

[e]Rating anchor: total score=0-34 (17 items).

Discussion

Principal Findings

This study aimed to delineate the psychometric properties of the PSC-17-Y in Spanish adolescents. As expected, this study found evidence of a 3-factor solution, as in the original English version, and also gave evidence of reliability and validity (structural, convergent, and criterion) to assess several psychopathology symptoms among adolescents. However, item 13 showed a lower loading (0.127) than the remaining items. This finding is consistent with Bergmann et al. [22], who validated the PSC-17-Y in English and reported a relatively low factor loading (0.233) for this item. Following the same logic as these authors, we maintained item 13 in the final set of items for the PSC-17-Y in Spanish, since it is important to keep the measure as simple as possible for respondents and clinicians to complete, score, and interpret and since the parent- and youth-reported short forms were identical except for this 1 item. We found that the inclusion of item 13 had a negligible impact on the psychometric properties of the PSC-17-Y in Spanish. As a result, we elected to add item 13 to the 16-item model and recommend a 17-item short form of the PSC-Y that uses the same 17 items on the same 3 subscales as the parent-reported PSC-17. Furthermore, we hypothesized that the problem with this item could be the wording and that future studies might review it to improve the saturation in the factor. The PSC-17-Y also showed an invariant structure across both genders, again consistent with Bergman et al [22].

Additionally, our study provided the first evidence, as far we know, for longitudinal invariance, indicating that the PSC-17-Y in Spanish adolescents is stable over time. This finding implies that it is reasonable to conclude that growth or development in observed scores over time can be attributed to actual development or changes in the construct under investigation, not measurement problems [48]. Further studies that replicate our study over more extended periods are, of course, needed.

Once gender and longitudinal invariance were tested, this study provided new evidence on gender-attributable differences in the PSC-Y scales. The gender differences found in this study were consistent with the overwhelming prior research establishing that females are more likely to express INT

symptoms and males are more likely to express EXT symptoms [49-51]. Furthermore, the small ES found in these gender differences is also consistent with other ones, which highlighted the small magnitude of gender differences in INT problems among children and adolescents [52]. These findings suggest that gender differences should be considered when pediatric and mental health professionals interpret PSC-17-Y results. Furthermore, gender is a crucial variable in the relationship between INT symptoms and suicide among adolescents, increasing this risk in females [53].

Regarding reliability evidence, our study showed McDonald ω values between .64 and .76, which are slightly lower than those shown by previous studies. Gardner et al [16] reported high internal consistency (.79 for INT symptoms, .83 for EXT symptoms, and .83 for ATT), and in the same order, Bergmann et al [22], with 16 items, found consistency values of .81, .74, and .69, respectively. The lowest internal consistency value was ATT (.65 and .64 for females and males, respectively). The removal of item 13 could improve the internal consistency of the ATT subscale slightly to values of .70 and .71 for females and males, respectively but at the cost of losing the original 17-item structure of the questionnaire.

Concerning other sources of convergent and criterion validity evidence of the PSC-17-Y, all 3 subscales correlated positively with the measure of distress and negatively with well-being, HRQoL, and social-emotional competencies, indicating a higher correlation between PSC-17-Y INT problems and the remaining convergent and criterion validity measures. This finding is consistent with previous studies on the PSC-17-Y, such as Parker et al [23], who reported screening validity of the PSC-17-Y in terms of higher scores on the PSC ATT and INT subscales among youth with any lifetime mental health diagnosis, as well as that ATT and INT subscale scores (but not EXT) are significantly correlated with psychosis, depression, and anxiety disorders. Thus, both our study and Parker et al's [23] provide support for the convergent and discriminant validity of the PSC-17-Y.

Finally, regarding percentile scores, score distributions showed a positive asymmetric distribution, but these normative data can help locate specific and general psychopathological problems among adolescents. Our rates would be suggestive of adolescents presenting scores compatible with mental health symptoms likely being in the clinical range. These data are equivalent to those reported in previous studies using different versions and cut-off scores for the PSC, PSC-17, and PSC-17-Y: 5%-25% of children were screened positive [15,16,54-56]. For example, a study using the 90th percentile score on the PSC found that 10.4% of children had problems on the OVR scale [54]. Additionally, the prevalence rates found in our study are consistent with wide international reviews on estimates of mental health prevalence among adolescents [1].

Limitations

The absence of an equivalent, well-established measure of INT, EXT, and ADHD measures; the absence of data on the area under the curve at optimal cut-off points in this study; and the sample's representativeness because we recruited the sample from the southeast of Spain exclusively were limitations of this study. In addition, there were only 2 waves of assessment in a short period of time (7 months), so it is recommended that future studies replicate the longitudinal invariance findings over longer intervals.

Conclusion

This study showed that the PSC-17-Y is a useful, reliable, and valid ultrabrief screening measure for detecting mental health problems in adolescents and can be administered over the internet. More specifically, this study provided evidence of the reliability and validity (structural and convergent-discriminant) of the Spanish version of the PSC-17-Y for adolescents.

Finally, these findings are significant for the scientific community. Therefore, this work has allowed us to extend the evidence of the validity of the PSC-17-Y to another language and country (Spain) in a large sample of adolescents, where scores were invariant over time irrespective of gender. This is a requirement that few instruments meet or for which evidence has been provided. All of this supports the reliability of the PSC-17-Y's assessments and its use in clinical contexts, such as monitoring the development of symptomatology. The fact that the PSC-17-Y is a tool that is easy to administer is another support for its use in clinical contexts.

Having instruments such as the PSC-17-Y with established reliability meets an especially important need during COVID-19 times, which have been characterized by an increase in mental health problems among children and adolescents [3-5] and a possibly greater need for case identification and outcome measurement.

Lastly, the results also support the use of the PSC-17-Y in longitudinal research, for example, for the study of the temporal trajectories of psychopathology in children, facilitating, among other things, reliability in the evaluation of the effectiveness of treatments. In addition to its usefulness in research, the PSC-17-Y is an instrument with applicability in the clinical setting, specifically in both primary care and specialty mental health units, as a screening tool for mental health problems in children and adolescents that is valid for monitoring changes in functioning over time.

Acknowledgments

We would like to thank collaboration and support by the Dirección General de Atención a la Diversidad y Calidad Educativa de la Consejería de Educación, Juventud y Deportes de la Región de Murcia (General Department for Attention to Diversity and Educational Quality of the Regional Ministry of Education, Youth and Sports of the Region of Murcia, Spain) and the Secretaría Autonómica de Educación e Investigación de la Conselleria de Educación, Investigación, Cultura y Deporte de la Generalitat Valenciana (Regional Secretariat of Education and Research of the Regional Ministry of Education, Research, Culture and Sport

of the Generalitat Valenciana, Spain), as well as to the educational centers, the authorization of parents and legal guardians, and especially the students, in the completion of the psychological assessment protocol, which made this study possible. We would also like to thank Javier Guzman (University of Desarrollo, Chile) for his assistance in adapting the PSC-17-Y into European Spanish.

This research was supported by a grant for I+D+i projects under award number PSI2017-88280-R (AEI/FEDER, UE) and 2 grants for Research Networks (PSI2015-70943-REDT and PSI2017-90650-REDT) from the Ministry of Economy, Industry and Competitiveness of the Spanish Government, as well as 3 predoctoral fellowships cofinanced by the European Social Fund and the Regional Ministry of Education, Research, Culture and Sport from the Generalitat Valenciana (RF: ACIF/2019/052; BM-A: ACIF/2019/055). VV-A is supported by a scholarship from the Universitat Jaume I (PREDOC/18/12); JMH and MM receive salary support from the Fuss Family Fund.

Authors' Contributions

JAP is principal investigator 1 of the funded project. He participated in the design of the psychological assessment protocol, collaborated in the conceptualization of this study by providing theoretical knowledge and professional and research background, and drafted the initial manuscript. VV-A participated in the conceptualization of the study by providing an analytical-methodological perspective, undertook the analysis of the data that made up the results of the study, and carried out a review of the statistical content. RF participated in the conceptualization of the study and in the design of the assessment protocol, coordinated and supervised the data collection, managed the database, reviewed the contents and terminology of the manuscript, and adapted its presentation format to the formatting requirements. BM-A participated in the design of the assessment protocol, coordinated and supervised the data collection, managed the online survey, collaborated in the selection and adaptation of the instruments, assisted in literature review tasks, and carried out a general review of contents. JCM is principal investigator 2 of the funded project. He participated in the design of the psychological assessment protocol, in the review and editing of the theoretical-methodological content, and in the adaptation of the manuscript to the formatting requirements. JH collaborated in the review of English editing and the final version of the manuscript. MM is the cocreator of the Pediatric Symptom Checklist (PSC). He consulted on the adaptation of the PSC into European Spanish and reviewed the final revision of the manuscript. All the authors approved the final manuscript, as submitted, and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar 03;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
2. Fonseca-Pedrero E, Ortuño-Sierra J, Pérez-Albéniz A. Emotional and behavioural difficulties and prosocial behaviour in adolescents: a latent profile analysis. Rev Psiquiatr Salud Ment (Engl Ed) 2020;13(4):202-212. [doi: 10.1016/j.rpsm.2020.01.002] [Medline: 32444209]
3. Francisco R, Pedro M, Delvecchio E, Espada JP, Morales A, Mazzeschi C, et al. Psychological symptoms and behavioral changes in children and adolescents during the early phase of COVID-19 quarantine in three European countries. Front Psychiatry 2020 Dec 3;11:570164 [FREE Full text] [doi: 10.3389/fpsyt.2020.570164] [Medline: 33343415]
4. Espada J, Orgilés M, Piqueras J, Morales A. Good practices in child and adolescent psychological care before COVID-19. Clínica y Salud 2020 2020 Jul 01;31(2):109-113. [doi: 10.5093/clysa2020a14]
5. Tamarit A, de la Barrera U, Monaco E, Schoeps K, Montoya-Castilla I. Psychological impact of COVID-19 pandemic in Spanish adolescents: risk and protective factors of emotional symptoms. Rev Psicol Clínica con Niños y Adolesc 2020 Sep 01;7(3):73-80. [doi: 10.21134/rpcna.2020.mon.2037]
6. World Health Organization. Global Accelerated Action for the Health of Adolescents (AA-HA!): Guidance to Support Country Implementation. Geneva: WHO; 2017:176.
7. Brown J, Wissow L. Screening to identify mental health problems in pediatric primary care: considerations for practice. Int J Psychiatry Med 2010 May 6;40(1):1-19. [doi: 10.2190/PM.40.1.a] [Medline: 20565041]
8. Wissow L, Brown J, Fothergill K, Gadomski A, Hacker K, Salmon P, et al. Universal mental health screening in pediatric primary care: a systematic review. J Am Acad Child Adolesc Psychiatry 2013 Nov;52(11):1134-1147.e23 [FREE Full text] [doi: 10.1016/j.jaac.2013.08.013] [Medline: 24157388]
9. Webb MJ, Kauer SD, Ozer EM, Haller DM, Sanci LA. Does screening for and intervening with multiple health compromising behaviours and mental health disorders amongst young people attending primary care improve health outcomes? A systematic review. BMC Fam Pract 2016 Aug 04;17:104 [FREE Full text] [doi: 10.1186/s12875-016-0504-1] [Medline: 27488823]

10. Connors E, Arora P, Blizzard A, Bower K, Coble K, Harrison J, et al. When behavioral health concerns present in pediatric primary care: factors influencing provider decision-making. J Behav Health Serv Res 2018 Jul;45(3):340-355 [FREE Full text] [doi: 10.1007/s11414-017-9580-9] [Medline: 29209899]
11. Ham P, Allen C. Adolescent health screening and counseling. Am Fam Physician 2012 Dec 15;86(12):1109-1116 [FREE Full text] [Medline: 23316983]
12. Jeffrey J, Klomhaus A, Enenbach M, Lester P, Krishna R. Self-report rating scales to guide measurement-based care in child and adolescent psychiatry. Child Adolesc Psychiatr Clin N Am 2020 Oct;29(4):601-629. [doi: 10.1016/j.chc.2020.06.002] [Medline: 32891365]
13. Jellinek M, Murphy JM. Psychosocial problems, screening, and the Pediatric Symptom Checklist. Dev Behav News 1999 Jan;8(1):6-10.
14. Jellinek MS, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. J Pediatr 1986 Aug 01;109(2):371-378. [doi: 10.1016/s0022-3476(86)80408-5] [Medline: 3734977]
15. Jellinek MS, Murphy JM, Robinson J, Feins A, Lamb S, Fenton T. Pediatric Symptom Checklist: screening school-age children for psychosocial dysfunction. J Pediatr 1988 Feb;112(2):201-209. [doi: 10.1016/s0022-3476(88)80056-8] [Medline: 3339501]
16. Gardner W, Murphy M, Childs G, Kelleher K, Pagano M, Jellinek M, et al. The PSC-17: a brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. Ambul Child Heal 1999;5(3):225-236.
17. Gardner W, Lucas A, Kolko DJ, Campo JV. Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. J Am Acad Child Adolesc Psychiatry 2007 May;46(5):611-618 [FREE Full text] [doi: 10.1097/chi.0b013e318032384b] [Medline: 17450052]
18. Murphy JM, Bergmann P, Chiang C, Sturner R, Howard B, Abel MR, et al. The PSC-17: subscale scores, reliability, and factor structure in a new national sample. Pediatrics 2016 Sep 12;138(3):e20160038 [FREE Full text] [doi: 10.1542/peds.2016-0038] [Medline: 27519444]
19. Pagano ME, Cassidy LJ, Little M, Murphy JM, Jellinek AMS. Identifying psychosocial dysfunction in school-age children: The Pediatric Symptom Checklist as a self-report measure. Psychol Sch 2000 Mar 01;37(2):91-106 [FREE Full text] [doi: 10.1002/(SICI)1520-6807(200003)37:2<91::AID-PITS1>3.0.CO;2-3] [Medline: 22328794]
20. Gall G, Pagano M, Desmond M, Perrin J, Murphy J. Utility of psychosocial screening at a school-based health center. J Sch Health 2000 Sep 01;70(7):292-298 [FREE Full text] [doi: 10.1111/j.1746-1561.2000.tb07254.x] [Medline: 10981284]
21. Murphy JM, Nguyen T, Lucke C, Chiang C, Plasencia N, Jellinek M. Adolescent self-screening for mental health problems; demonstration of an internet-based approach. Acad Pediatr 2018;18(1):59-65. [doi: 10.1016/j.acap.2017.08.013] [Medline: 28870650]
22. Bergmann P, Lucke C, Nguyen T, Jellinek M, Murphy JM. Identification and utility of a short form of the Pediatric Symptom Checklist-Youth Self-Report (PSC-17-Y). Eur J Psychol Assess 2020 Jan;36(1):56-64. [doi: 10.1027/1015-5759/a000486]
23. Parker EM, Jacobson J, Pullmann MD, Kerns SEU. Identifying psychosocial problems among children and youth in the child welfare system using the PSC-17: exploring convergent and discriminant validity with multiple informants. Child Psychiatry Hum Dev 2019 Feb;50(1):108-120. [doi: 10.1007/s10578-018-0824-6] [Medline: 29961167]
24. Muñiz J, Elosua P, Hambleton RK, International Test Commission. International Test Commission Guidelines for test translation and adaptation: second edition. Psicothema 2013;25(2):151-157. [doi: 10.7334/psicothema2013.24] [Medline: 23628527]
25. Guzmán J, Kessler RC, Squicciarini AM, George M, Baer L, Canenguez KM, et al. Evidence for the effectiveness of a national school-based mental health program in Chile. J Am Acad Child Adolesc Psychiatry 2015 Oct;54(10):799-807.e1. [doi: 10.1016/j.jaac.2015.07.005] [Medline: 26407489]
26. Leiva L, Rojas R, Peña F, Vargas B, Scquicciarini A. Detectando las dificultades emocionales y conductuales en la escuela: validación de PSC-17. REV IBEROAM DIAGN EV 2019 Jan;50(4):95-105 [FREE Full text] [doi: 10.21865/ridep50.1.08]
27. Dowdy E, Furlong MJ, Nylund-Gibson K, Moore S, Moffa K. Initial validation of the Social Emotional Distress Survey–Secondary to support complete mental health screening. Assess Eff Interv 2018 Jan 05;43(4):241-248. [doi: 10.1177/1534508417749871]
28. Keyes CLM, Wissing M, Potgieter JP, Temane M, Kruger A, van Rooy S. Evaluation of the mental health continuum-short form (MHC-SF) in setswana-speaking South Africans. Clin Psychol Psychother 2008;15(3):181-192. [doi: 10.1002/cpp.572] [Medline: 19115439]
29. Reinhardt M, Horváth Z, Morgan A, Kökönyei G. Well-being profiles in adolescence: psychometric properties and latent profile analysis of the mental health continuum model: a methodological study. Health Qual Life Outcomes 2020 Apr 06;18(1):95 [FREE Full text] [doi: 10.1186/s12955-020-01332-0] [Medline: 32252785]
30. Piqueras JA, Vidal-Arenas V, Falco R, Moreno-Amador B, Marzo JC, Keyes C. Validation of the Mental Health Continuum-Short Form (MHC-SF) for multidimensional assessment of subjective well-being in Spanish adolescents. Psicothema 2021 Nov 18:1-15 (forthcoming).
31. Ravens-Sieberer U, Erhart M, Rajmil L, Herdman M, Auquier P, Bruil J, European KIDSCREEN Group. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and

health-related quality of life. Qual Life Res 2010 Dec;19(10):1487-1500 [FREE Full text] [doi: 10.1007/s11136-010-9706-5] [Medline: 20668950]
32. Furlong MJ, You S, Renshaw TL, Smith DC, O'Malley MD. Preliminary development and validation of the Social and Emotional Health Survey for secondary school students. Soc Indic Res 2013 Jun 27;117(3):1011-1032. [doi: 10.1007/s11205-013-0373-0]
33. Piqueras JA, Rodriguez-Jimenez T, Marzo JC, Rivera-Riquelme M, Martinez-Gonzalez AE, Falco R, et al. Social Emotional Health Survey-Secondary (SEHS-S): a universal screening measure of social-emotional strengths for Spanish-speaking adolescents. Int J Environ Res Public Health 2019 Dec 07;16(24):4982 [FREE Full text] [doi: 10.3390/ijerph16244982] [Medline: 31817874]
34. Ato M, López-García JJ, Benavente A. Un sistema de clasificación de los diseños de investigación en psicología. Anal Psicol 2013 Oct 01;29(3):1038-1059. [doi: 10.6018/analesps.29.3.178511]
35. Montero I, León O. A guide for naming research studies in psychology. Int J Clin Heal Psychol 2007;7(3):847-862 [FREE Full text]
36. Li C. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods 2016 Sep;48(3):936-949. [doi: 10.3758/s13428-015-0619-7] [Medline: 26174714]
37. Marsh HW, Hau K, Wen Z. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Struct Equ Model 2004 Jul;11(3):320-341. [doi: 10.1207/s15328007sem1103_2]
38. Weston R, Gore PA. A brief guide to structural equation modeling. Couns Psychol 2016 Jun 30;34(5):719-751. [doi: 10.1177/0011000006286345]
39. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 2002 Apr;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
40. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model 2007 Dec 05;14(3):464-504. [doi: 10.1080/10705510701301834]
41. Dimitrov D. Statistical Methods for Validation of Assessment Scale Data in Counseling and Related Fields. Alexandria, VA: American Counseling Association; 2012.
42. Cohen J. Statistical Power Analysis for the Behavioural Sciences. Hillside, NJ: Lawrence Earlbaum Associates; 1988.
43. McDonald R. Test Theory: A Unified Treatment. Hillside, NJ: Lawrence Erlbaum Associates; 1999.
44. 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]
45. Lipsey M, Wilson D. Practical Meta-Analysis. Thousand Oaks, CA: Sage; 2001.
46. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin J. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 2014 Dec 09;12:176 [FREE Full text] [doi: 10.1186/s12955-014-0176-2] [Medline: 25492701]
47. Tabachnick B, Fidell L, Ullman J. Using Multivariate Statistics (7th Ed.). Boston, MA: Pearson; 2019.
48. Millsap R, Cham H. Investigating factorial invariance in longitudinal data. In: Laursen B, Little TD, Card NA, editors. Handbook of Developmental Research Methods. New York, NY: Guilford Press; 2012:109-126.
49. Jacobson JH, Pullmann MD, Parker EM, Kerns SEU. Measurement based care in child welfare-involved children and youth: reliability and validity of the PSC-17. Child Psychiatry Hum Dev 2019 Apr 27;50(2):332-345. [doi: 10.1007/s10578-018-0845-1] [Medline: 30264230]
50. Lewis AJ, Sae-Koew JH, Toumbourou JW, Rowland B. Gender differences in trajectories of depressive symptoms across childhood and adolescence: a multi-group growth mixture model. J Affect Disord 2020 Jan 01;260:463-472. [doi: 10.1016/j.jad.2019.09.027] [Medline: 31539681]
51. Rocchino GH, Dever BV, Telesford A, Fletcher K. Internalizing and externalizing in adolescence: the roles of academic self-efficacy and gender. Psychol Sch 2017 Aug 06;54(9):905-917. [doi: 10.1002/pits.22045]
52. Piqueras JA, Garcia-Olcina M, Rivera-Riquelme M, Martinez-Gonzalez AE, Cuijpers P. DetectaWeb-Distress Scale: a global and multidimensional web-based screener for emotional disorder symptoms in children and adolescents. Front Psychol 2021 Feb 15;12:627604 [FREE Full text] [doi: 10.3389/fpsyg.2021.627604] [Medline: 33658965]
53. Soto-Sanz V, Castellví P, Piqueras JA, Rodríguez-Marín J, Rodríguez-Jiménez T, Miranda-Mendizábal A, et al. Internalizing and externalizing symptoms and suicidal behaviour in young people: a systematic review and meta-analysis of longitudinal studies. Acta Psychiatr Scand 2019 Jul;140(1):5-19. [doi: 10.1111/acps.13036] [Medline: 30980525]
54. Bala G, Golubović S, Milutinović D, Katić R. Psychometric properties of the Pediatric Symptom Checklist in preschool children in Serbia. Med Glas (Zenica) 2012 Aug;9(2):356-362. [Medline: 22926377]
55. Reijneveld SA, Vogels AG, Hoekstra F, Crone MR. Use of the Pediatric Symptom Checklist for the detection of psychosocial problems in preventive child healthcare. BMC Public Health 2006 Jul 27;6(1):197 [FREE Full text] [doi: 10.1186/1471-2458-6-197] [Medline: 16872535]
56. Lowenthal E, Lawler K, Harari N, Moamogwe L, Masunge J, Masedi M, et al. Validation of the Pediatric Symptom Checklist in HIV-infected Batswana. J Child Adolesc Ment Health 2011 Jan 01;23(1):17-28 [FREE Full text] [doi: 10.2989/17280583.2011.594245] [Medline: 22685483]

Abbreviations

ADHD: attention deficit hyperactivity disorder
ATT: attention deficit hyperactivity disorder symptoms
CFA: confirmatory factor analysis
CFI: comparative fit index
ES: effect size
EXT: externalizing
HRQoL: health-related quality of life
INT: internalizing
MHC-SF: Mental Health Continuum-Short Form
OVR: overall
PSC-17-Y: Short Form of the Pediatric Symptom Checklist-Youth Self-Report
PSC-17: Pediatric Symptom Checklist-Parent Version
PSC: Pediatric Symptom Checklist
RMSEA: root-mean-square error of approximation
SEDS-S: Social-Emotional Distress Survey-Secondary
SEHS-S: Social-Emotional Health Survey-Secondary
TLI: Tucker-Lewis index
WLSMV: weighted least squares means and variance adjusted

Edited by R Kukafka; submitted 10.06.21; peer-reviewed by A Tannoubi, M Salimi; comments to author 30.07.21; revised version received 06.10.21; accepted 29.10.21; published 01.12.21.

Please cite as:
Piqueras JA, Vidal-Arenas V, Falcó R, Moreno-Amador B, Marzo JC, Holcomb JM, Murphy M
Short Form of the Pediatric Symptom Checklist-Youth Self-Report (PSC-17-Y): Spanish Validation Study
J Med Internet Res 2021;23(12):e31127
URL: https://www.jmir.org/2021/12/e31127
doi:10.2196/31127
PMID:34855614

Original Paper

Assessing Cognitive Function in Multiple Sclerosis With Digital Tools: Observational Study

Wan-Yu Hsu[1*], PhD; William Rowles[1*], BA; Joaquin A Anguera[1,2,3], PhD; Annika Anderson[1], BA; Jessica W Younger[1,2], PhD; Samuel Friedman[1], BA; Adam Gazzaley[1,2,3,4], MD, PhD; Riley Bove[1], MD

[1]Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, CA, United States

[2]Neuroscape, University of California, San Francisco, CA, United States

[3]Department of Psychiatry, University of California, San Francisco, CA, United States

[4]Department of Physiology, University of California, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Riley Bove, MD
Department of Neurology
Weill Institute for Neurosciences
University of California
675 Nelson Rising Lane
San Francisco, CA, 94158
United States
Phone: 1 415 595 2795
Email: Riley.Bove@ucsf.edu

Abstract

Background: Cognitive impairment (CI) is one of the most prevalent symptoms of multiple sclerosis (MS). However, it is difficult to include cognitive assessment as part of MS standard care since the comprehensive neuropsychological examinations are usually time-consuming and extensive.

Objective: To improve access to CI assessment, we evaluated the feasibility and potential assessment sensitivity of a tablet-based cognitive battery in patients with MS.

Methods: In total, 53 participants with MS (24 [45%] with CI and 29 [55%] without CI) and 24 non-MS participants were assessed with a tablet-based cognitive battery (Adaptive Cognitive Evaluation [ACE]) and standard cognitive measures, including the Symbol Digit Modalities Test (SDMT) and the Paced Auditory Serial Addition Test (PASAT). Associations between performance in ACE and the SDMT/PASAT were explored, with group comparisons to evaluate whether ACE modules can capture group-level differences.

Results: Correlations between performance in ACE and the SDMT (R=–0.57, *P*<.001), as well as PASAT (R=–0.39, *P*=.01), were observed. Compared to non-MS and non-CI MS groups, the CI MS group showed a slower reaction time (CI MS vs non-MS: *P*<.001; CI MS vs non-CI MS: *P*=.004) and a higher attention cost (CI MS vs non-MS: *P*=.02; CI MS vs non-CI MS: *P*<.001).

Conclusions: These results provide preliminary evidence that ACE, a tablet-based cognitive assessment battery, provides modules that could potentially serve as a digital cognitive assessment for people with MS.

Trial Registration: ClinicalTrials.gov NCT03569618; https://clinicaltrials.gov/ct2/show/NCT03569618

(J Med Internet Res 2021;23(12):e25748) doi:10.2196/25748

KEYWORDS

cognition; digital health; mHealth; multiple sclerosis; cognitive assessment

Introduction

Background

Multiple sclerosis (MS) is a chronic inflammatory and neurodegenerative disorder, and it is the leading cause of major disability in young adults. Cognitive impairment (CI) occurs in 30%-70% of patients with MS [1,2], even in the absence of physical impairment [1,3,4]. CI is one of the most debilitating manifestations of MS and can have a profound influence on a patient's personal independence and quality of life, interfering with social functioning and employment. Since 2014, CI assessment has been one of the measure specifications of the American Academy of Neurology's MS Quality Measurement Set [5]. Ideally, patients with MS should undergo a complete cognitive assessment and routinely repeat the examination to detect cognitive changes overtime and to start timely treatment, if needed. However, to date, cognitive assessment in MS relies on a comprehensive neuropsychological examination, which is time-consuming and extensive; therefore, it makes it difficult to include cognitive assessment/monitoring as part of MS standard care.

Cognitive Assessment with Digital Tools

Integrating digital tools into clinical settings can reduce the time and cost associated with cognitive examination and further allows repeated assessments, which provide more precise monitoring of cognitive performance and longitudinal changes. One more feature of digital tools is the capability of remote and self-administration, which can relieve the burden of travel to the clinic, due to deficits in mobility or cognition for many patients. Moreover, with remote administration features, data collection can be performed in a nonclinical, real-life setting, which allows sampling of cognitive performance more closely reflecting real-world cognitive function [6]. The self-administered feature also reduces common stressors for patients who get nervous during structured testing in clinical settings.

With advanced technology, remote, computerized platforms for cognitive assessment and treatment, using personalizing features, including adaptive staircase algorithms for populations with cognitive deficits [7-14], have been developed. In Alzheimer's disease–related dementias, digital cognitive assessment tools have shown reliability in measuring longitudinal cognitive changes in individuals with no CI, mild CI, and dementia [15,16] and have exhibited cross-sectional sensitivity to cerebrospinal fluid amyloid-ß levels [17]. In schizophrenia, digital assessments have also shown effectiveness in identifying deficits across different cognitive domains [18,19]. Moreover, tablet-based cognitive assessment has been validated to differentiate cognitive control ability between children with and without 16p11.2 deletion, a genetic variation implicated in attention deficit/hyperactivity disorder and autism [8]. Although the development and validation of digital cognitive assessment tools have been growing, the investigations of remote, digital CI assessments in MS have been scarce [20-23]; therefore, there is a need to deepen the exploration of digital cognitive evaluation for improving access to CI assessments in order to thereby navigate problems related to cognitive issues and further reduce the impact of CI on patients' lives.

Aims and Overview of the Study

The goal of this study was to evaluate the feasibility and potential assessment sensitivity of a tablet-based cognitive assessment battery in patients with MS, focusing on the most commonly affected cognitive domains in MS: processing speed, attention, executive function, and memory [1,2]. To accomplish this, a tablet-based cognitive assessment battery (Adaptive Cognitive Evaluation [ACE]; see the Methods section) that measures different aspects of high-order cognitive function (eg, attention, working memory, speed of information processing, and executive function) [24], was tested in 53 participants with MS and 24 participants without MS. ACE was developed by Neuroscape at the University of California, San Francisco (UCSF) [24]. It has a user-friendly interface as well as adaptive algorithms, which modulate the challenge level of a task on a trial-by-trial basis based on individual performance. In this study, 3 modules (Boxed, Sustained Attention ACE Task [SAAT], Spatial Span) assessing different aspects of cognitive control ability and 1 module (Basic Reaction Time [BRT]) measuring the basic response speed were included (see the Methods section) for a preliminary examination of the construct validity of the ACE battery. The Symbol Digit Modalities Test (SDMT), a test considered the most sensitive measurement for the evaluation of cognitive involvement and information processing speed in the early MS course [25,26], was also administered. Given that information processing speed has been shown to account for impairments in high-level cognitive functions in MS [27-30], the relationship between performance in the SDMT and ACE modules was investigated. To further delineate whether ACE modules can differentiate different levels of cognitive function, performance differences among participants with MS with and without CI, as well as participants without MS, were examined.

We hypothesized that there would be a correlation between the SDMT score and ACE performance in accordance with the relative consequence theory of information processing speed [30-32], in which impaired processing speed is considered the key deficit underlying CI in MS [27-30]. In addition, as a tool for cognitive assessment in MS, ACE would reveal group-level differences between participants with MS with and without CI, as well as participants without MS.

Methods

Participants

In total, 53 adults with clinically definite MS [33], mean age 51.8 (1.7) years, were recruited from the University of California, San Francisco Multiple Sclerosis and Neuroinflammation Center between April 2018 and January 2019 with the following inclusion criteria: internet connection available at home or in the work environment and free of relapses or steroid use in the past month. Patients with severe visual, cognitive, or motor impairment that would preclude the use of a tablet-based tool were excluded. A group of 24 adults without MS (non-MS), mean age 46.0 (3.7) years, with no chronic autoimmune diseases were also recruited from the UCSF

staff, willing family members of patients in the clinic, and other eligible and willing volunteers.

Standard Approval, Registration, and Patient Consent

All procedures performed in the study involving human participants were approved by the Committee for Human Research at the UCSF. Written informed consent was obtained from each participant. The trial is registered with clinicaltrials.gov (NCT03569618).

Study Design

To evaluate ACE, both ACE and the SDMT were administered to all participants, including 53 adults with MS and 24 adults without MS (non-MS). All participants with MS were recruited as part of studies to determine the feasibility [20] and preliminary efficacy [34] of a digital cognitive treatment, as previously described [34]. As part of a published study in which another tablet-based assessment (ie, EVO Monitor) was investigated [35], this study contains data that have not been analyzed or published in a larger trial (clinicaltrials.gov NCT03569618). The analysis of this study was based on baseline performance data (ie, before any cognitive intervention) of our feasibility [20] and efficacy [34] trials, where participants underwent cognitive testing, including ACE and the SDMT. The 2-hour baseline session began with standard measures (SDMT, the Paced Auditory Serial Addition Test [PASAT], the California Verbal Learning Test Second Edition, and the Brief Visuospatial Memory Test Revised), followed by digital cognitive assessment with ACE (BRT, Boxed, SAAT, Spatial Span) and EVO Monitor (data presented in [35]). Standard measures were administered by a study coordinator. Digital tool assessment was self-guided; however, a study coordinator sat in with the participants to answer questions and clarify aspects of the directions, if needed. The baseline session did not include any predetermined break, while participants were informed at the beginning of the visit that they could take a break at any time, if needed. Task order was predetermined (as described above) and remained consistent through the whole study. Only the SDMT and PASAT were included in the analysis of this study, given that they are the most widely used standard measures for people with MS and the cognitive domains being evaluated by these tests are close to cognitive aspects that ACE is designed to test for (ie, attention, working memory, speed of information processing, and executive function).

Cognitive Measures

Standard Measures: SDMT and PASAT

SDMT is a widely used measure of selective attention and information processing speed in MS [25,26], which requires the participant to substitute geometric symbols for numbers while scanning a response key. The participants were presented with a page headed by a key that pairs 9 symbols with the single digits 1-9. Rows below showed only symbols, and the task was to write the correct number in the spaces below based on the key row. After finishing the first 10 items with guidance, correct responses being made within 90 seconds were counted as the SDMT score.

Since PASAT has also been used extensively to test information processing speed, attention, and working memory in MS [36], we included it as a standard measure for participants with MS recruited in this study. The participants were instructed to listen to numbers presented every 3 seconds and add the number they hear to the number they heard before (rather than giving a running total). The PASAT score was defined as the total number of correct answers out of 60 possible answers.

Digital Cognitive Assessment Battery: ACE

Tasks within ACE (Figure 1) followed a similar schematic: across modules, the probe or target (as specified in the individual module description below) was displayed either until a response was made or until the maximum reaction time (RT) limit was reached. After each trial, the trial-level feedback, either a green (correct response was made within the RT limit), yellow (correct response was made outside of the RT limit), or red (incorrect response) centralized fixation cross was displayed for 200 ms, followed by a standard 1000 ms intertrial interval.

The *BRT* task was designed to index the basic response speed of participants on a simple task with minimal loading on executive function skills [37]. Participants were instructed to press a button at the bottom of the screen as fast as they could when they detected a symbol (target) that appeared in the center of the screen. The target always appeared without distraction. The BRTs were measured for both index fingers. Participants first completed 5 practice trials for each of their right and left index fingers, followed by 20 experimental trials per index finger. This task started with a maximum RT limit of 500 ms and a response window of 500 ms that adapted for each participant according to trialwise performance. An average RT was measured for each hand. Only data from the dominant hand were included as each participant's BRT in the following analyses.

The *Boxed* task was designed to measure visual search performance across different types and number of distractors [38]. Participants were presented with an array of either 4 or 12 Landolt squares (ie, squares with gaps) with an opening on 1 side until participants located the target (a green box with a gap on the top or bottom) and indicated the location of the gap (top or bottom) by tapping with their dominant hand on a button with either "top" or "bottom." Two distinct search modes were included: a feature search, where red Landolt squares were present in addition to the single green target, allowing the target to be located based solely on object color, and a conjunction search, where distractor boxes were green and red, with all green distractor boxes having gaps on either side and red distractor boxes having gaps on the top and bottom, similar to the green target box. As such, participants had to search the array for the target square based on a conjunction of features: both color and position of the opening. In addition to the 2 search types (feature and conjunction), there were 2 distractor load conditions: a low load with 1 target and 3 distractors and a high load with 1 target and 11 distractors. Participants completed 8 practice trials of each condition (32 total) before moving on to the experimental task with 25 trials per condition (100 total). This task started with a maximum RT limit of 1500 ms and a response window of 1000 ms that adapted for each participant according to

trialwise performance. Only correct responses made within the participants' adaptive response window were included in analysis. Task performance was assessed by examining the mean RT to correct responses for all trial types, including 4- and 12-item trials collapsed across both feature and conjunction search modes. The RT cost between target identification for feature and conjunction trials across each set size was measured as distraction cost = 12-item (conjunction and feature) – 4-item (conjunction and feature).

Figure 1. Screenshot of ACE. ACE: adaptive cognitive evaluation.

SAAT was developed based on the Tests of Variables of Attention (TOVA) [39] and was designed to include blocks that separately measure sustained attention (to an infrequent target) and inhibitory control (inhibiting a prepotent response to salient distractors). During a trial, a symbol (target) appeared at the top or bottom of the screen. Participants were instructed to press a button with the index finger of only their dominant hand when the target appeared at the top of the screen and to ignore the symbol and withhold a response when it appeared at the bottom. This task proceeded in 2 blocks. In the inhibitory control block, targets appeared on 27 of 40 (67%) of trials and required participants to withhold a (highly primed) response when a distractor appeared. In the sustained attention block, the target appeared on only 13 of 40 (33%) of trials, requiring participants to maintain attention to avoid missing an infrequent target. Participants completed 10 practice trials (6 target and 4 nontarget trials) and then 80 experimental trials, 40 in each block. This task started with a maximum RT limit of 600 ms and a response window of 600 ms that adapted for each participant according to trialwise performance. However, to avoid creating an artificially low response window, correct rejections did not affect the response window. Trials where no response was given (when a response was expected) or that were anticipatory (RT < 150 ms) were excluded from analyses. All remaining trials were evaluated for accuracy regardless of whether the response was within the response window. Thus, trials were only considered incorrect if an incorrect response was made (and not if they were correct but late). The mean RT to correct responses collapsed across block types were measured as task performance.

The *Spatial Span Task* is a computerized version of the Corsi Block-Tapping Test [40], which has frequently been used to assess visuospatial working memory capacity. On each trial, participants viewed a test array of 20 black circles that were cued sequentially in line with the typical administration of the Corsi Block-Tapping Task stimuli. Cued circles were lit in green, one at a time, sequentially. After the sequence of circles was complete and no longer displayed, participants were instructed to recall the location of each cued circle in the order they were shown and indicate the location and sequence by tapping each cued location in the cued order. Participants started with between 2 and 4 practice trials with 3-location sequences (ie, 3 cued circles). Participants practiced until 2 consecutive trials were answered incorrectly. Regardless of practice performance, participants then began the experimental task with a 3-location sequence. Once the participant completed 2 consecutive trials of the previous level without an error, they would advance to the next level that included an additional cued

circle, increasing the difficulty level. Participants completed as many levels as possible until 2 consecutive incorrect trials, at which point the task ended. Participants had unlimited time to respond for this task. Participants needed to have successfully completed at least 2 3-location sequence trials to be included in analysis. The highest level (ie, maximum number of items in a sequence) of the successful trial for each participant was defined as the spatial span, the measurement of working memory capacity.

Statistical Analysis

All numerical data are presented as the mean (SE). To evaluate the digital cognitive assessment battery (ie, ACE), Pearson correlation analyses were conducted to scrutinize the relationship between performance in the SDMT/PASAT and ACE modules. Partial correlation analyses with age, sex, years of education, and the BRT as covariates were applied, when appropriate. To examine whether the selected ACE modules can differentiate CI and non-CI participants with MS, participants with MS were divided into 2 subgroups (ie, CI and non-CI) according to their baseline SDMT z scores. Participants with an SDMT z score of <–1 based on published normative data [41] were characterized as CI. Differences between CI and non-CI participants with MS, as well as non-MS participants in terms of ACE performance, were examined by one-way ANOVA, with the BRT as a covariate to control for potential motor speed deficit in participants with MS. Two-tailed Student *t* tests were carried out for post hoc comparisons, when appropriate. The statistical significance threshold was set as $P \leqq .05$.

Results

Participants

A total of 53 participants with MS (mean age 51.8 [1.7] years) and 24 participants without MS (mean age 46.0 [3.7] years) completed the assessments; their demographic and clinical characteristics are summarized in Table 1. One-way ANOVA revealed no age differences between the groups ($F(2,76)=1.42$, $P=.24$). For categorical variables (ie, sex and race), chi-square tests showed no statistically significant association between groups and sex ($X^2(2)=5.31$, $P=.07$) as well as race ($X^2(4)=2.90$, $P=.59$). For analysis purposes, the 53 participants with MS were divided into CI (n=24 [45%]) and non-CI (n=29 [55%]) subgroups based on to their baseline SDMT z score. Figure 2 details the task completion rate.

Table 1. Demographic and clinical characteristics of participants.

Characteristics	MS[a]		Non-MS (n=24)
	CI[b] (n=24)	Non-CI (n=29)	
Age (years), mean (SE)	50.87 (2.51)	52.68 (2.35)	46.04 (3.72)
Sex (female), n (%)	17 (70)	23 (79)	12 (50)
Education (years), mean (SE)	16.50 (0.47)	16.79 (0.51)	16.16 (0.41)
Right-handedness, n (%)	22 (91)	23 (80)	50 (100)
Part- or full-time employed, n (%)	11 (45%)	15 (51%)	17 (70%)
SDMT[c] score, mean (SE)	34.79 (1.34)	49.34 (1.12)[d]	51.20 (2.65)[e]
SDMT z score, mean (SE)	–1.58 (0.08)	–0.05 (0.10)[d]	0.26 (0.20)[e]
Expanded Disability Status Scale (EDSS), median (IQR)	4 (1.75)	3 (1.5)	N/A[f]
Disease duration (years), mean (SE)	11.95 (1.83)	13.71 (1.49)	N/A
Race, n (%)			
White	22 (92%)	22 (76%)	19 (79%)
Black/African American	N/A	2 (7%)	1 (4%)
Other/unknown	2 (8%)	5 (17%)	4 (17%)
MS subtype, n (%)			
Relapsing-remitting	18 (75%)	23 (79%)	N/A
Primary progressive	2 (8%)	2 (7%)	N/A
Secondary progressive	3 (13%)	3 (10%)	N/A
Clinically isolated syndrome (CIS)	N/A	1 (3%)	N/A
Unknown	1 (4%)	N/A	N/A

[a]MS: multiple sclerosis.

[b]CI: cognitive impairment.

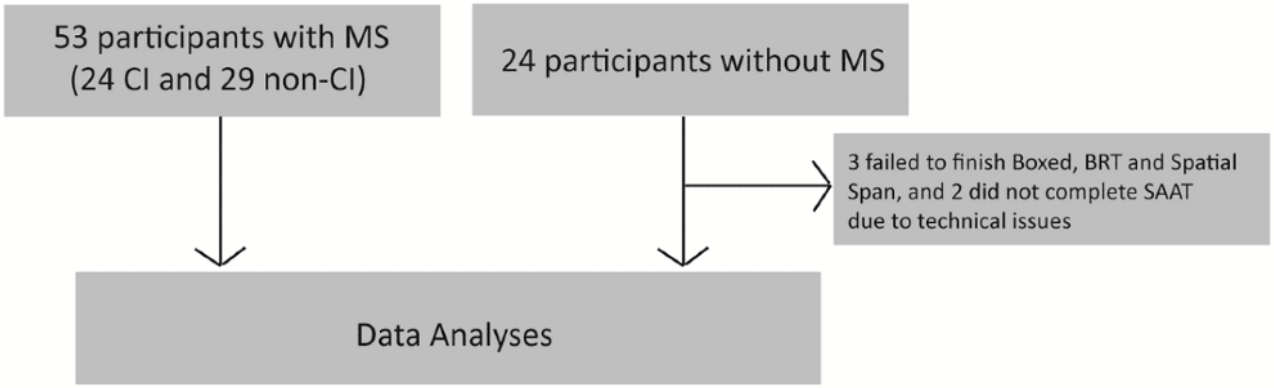
[c]SDMT: Symbol Digit Modalities Test.

[d]*P*<.001 for the comparison between CI and non-CI groups.

[e]*P*<.001 for the comparison between CI and non-MS groups.

[f]N/A: not applicable.

Figure 2. Task completion rate. ACE: Adaptive Cognitive Evaluation; BRT: basic reaction time; CI: cognitive impairment; MS: multiple sclerosis; SAAT: Sustained Attention ACE Task.



Correlation Between Standard Measures and ACE

To delineate associations between performance in standard measures (ie, SDMT and PASAT scores) and the tested digital cognitive platform (ie, ACE), Pearson correlation analyses were performed. The SDMT showed significant correlations with several ACE measures (Boxed RT: R=–0.57, *P*<.001; Boxed distraction cost: R=–0.28, *P*=.02; SAAT RT: R=–0.36, *P*=.001; Spatial Span: R=0.34, *P*=.003; Figure 3). Since the Boxed RT showed the strongest correlation with the SDMT, we further

performed an exploratory linear regression analysis to examine to what extent the Boxed RT value can be used to predict the SDMT score. The analysis revealed a moderate R-squared value of 0.333 with the regression equation SDMT = 82.55 – 0.038 × Boxed RT; 33% of the total variation in the SDMT score can be explained by the Boxed RT.

When controlling for age, sex, years of education, and the BRT with partial correlations, similar results were observed (Boxed RT: R=–0.44, *P*<.001; Boxed distraction cost: R=–0.28, *P*=.01; SAAT RT: R=–0.17, *P*=.15; Spatial Span: R=0.18, *P*=.12; Table 2). When we restricted the analyses to only participants with MS, SDMT correlations with the Boxed RT and Boxed distraction cost remained statistically significant (Boxed RT: R=–0.50, *P*<.001; Boxed distraction cost: R=–0.34, *P*=.01; SAAT RT: R=–0.22, *P*=.10; Spatial Span: R=0.24, *P*=.07). Again, in adjusted correlation analyses, we saw similar results (Boxed RT: R=–0.43, *P*=.002; Boxed distraction cost: R=–0.38, *P*=.01; SAAT RT: R=–0.16, *P*=.26; Spatial Span: R=0.18, *P*=.21; Table 2).

PASAT, which has also been extensively used to test cognitive function in MS, was tested in the participants with MS and also showed significant correlations with the Boxed RT (R=–0.39, *P*=.01) and Spatial Span (R=0.29, *P*=.03; Figure 4 and Table 3). These correlations remained significant after accounting for age, sex, years of education, and the BRT as covariates (Boxed RT: R=–0.40, *P*=.01; Spatial Span: R=0.34, *P*=.02). These results support the hypothesis that there is a correlational association between standard MS information processing measures and ACE measures.

Figure 3. Correlation between SDMT score and performance in ACE modules. ACE: Adaptive Cognitive Evaluation; RT: reaction time; SAAT: Sustained Attention ACE Task; SDMT: Symbol Digit Modalities Test.

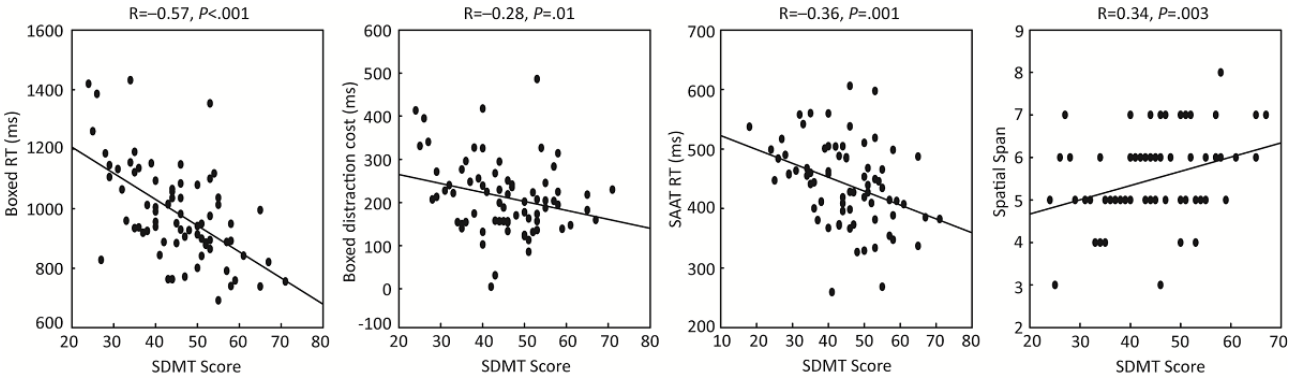


Table 2. Results of Pearson correlation analyses between SDMT[a] and ACE[b] measures.

ACE measures	Covariates	R	*P* value
All participants (N=77)			
Boxed RT[c]	N/A[d]	–0.57	*<.001* [e]
Boxed distraction cost	N/A	–0.28	*.01* [e]
SAAT[f] RT	N/A	–0.36	*.001* [e]
Spatial Span	N/A	0.34	*.003* [e]
Boxed RT	age, sex, edu[g], and BRT[h]	–0.44	*<.001* [e]
Boxed distraction cost	age, sex, edu, and BRT	–0.28	*.01* [e]
SAAT RT	age, sex, edu, and BRT	–0.17	.15
Spatial Span	age, sex, edu, and BRT	0.18	.12
Participants with MS[i] (N=53)			
Boxed RT	N/A	–0.50	*<.001* [e]
Boxed distraction cost	N/A	–0.34	*.01* [e]
SAAT RT	N/A	–0.22	.10
Spatial Span	N/A	0.24	.07
Boxed RT	age, sex, edu, and BRT	–0.43	*.002* [e]
Boxed distraction cost	age, sex, edu, and BRT	–0.38	*.007* [e]
SAAT RT	age, sex, edu, and BRT	–0.16	.26
Spatial Span	age, sex, edu, and BRT	0.18	.21

[a]SDMT: Symbol Digit Modalities Test.
[b]ACE: Adaptive Cognitive Evaluation.
[c]RT: reaction time.
[d]N/A: not applicable.
[e]*P* values in italic are significant.
[f]SAAT: Sustained Attention ACE Task.
[g]edu: years of education.
[h]BRT: basic reaction time.
[i]MS: multiple sclerosis.

Figure 4. Correlation between PASAT score and performance in ACE modules. ACE: Adaptive Cognitive Evaluation; PASAT: Paced Auditory Serial Addition Test; RT: reaction time.

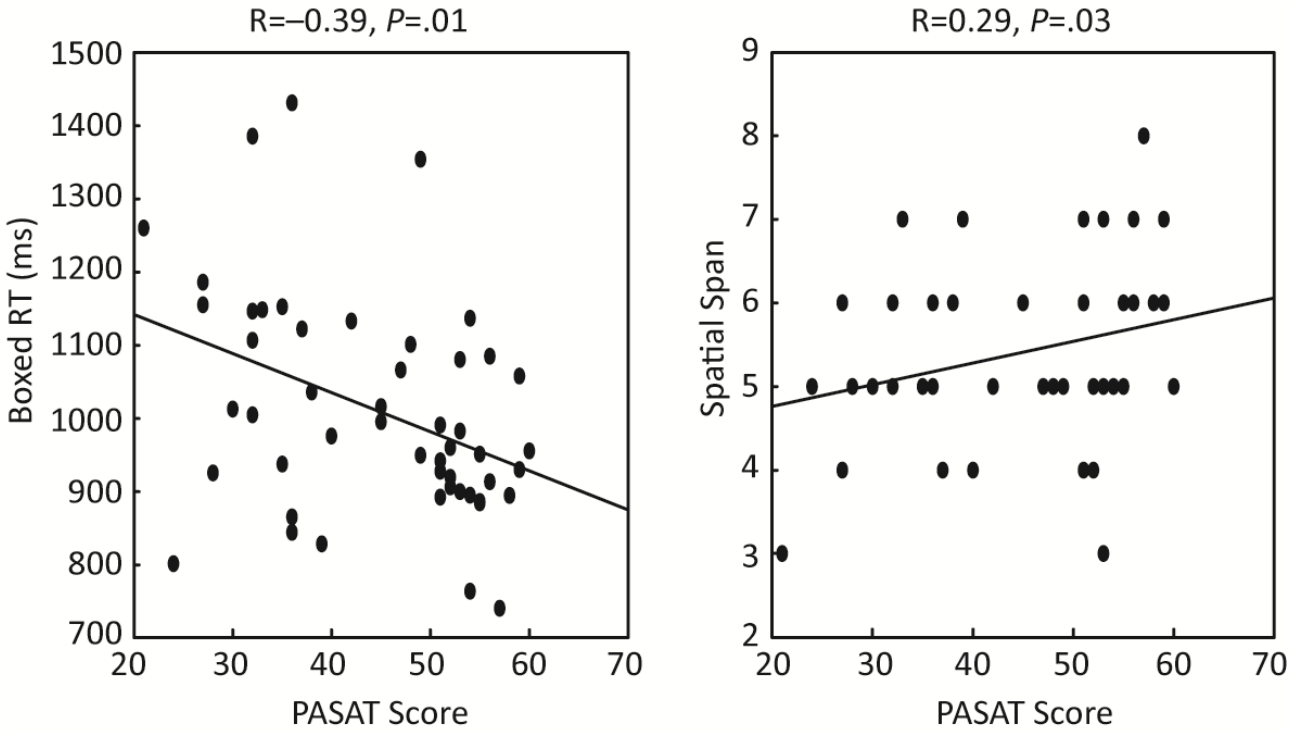


Table 3. Results of Pearson correlation analyses between PASAT[a] and ACE[b] measures in participants with MS[c] (N=53).

ACE measures	Covariates	R	*P* value
Boxed RT[d]	N/A[e]	–0.39	*.01* [f]
Boxed distraction cost	N/A	–0.25	.07
SAAT[g] RT	N/A	–0.03	.83
Spatial Span	N/A	0.29	*.03* [f]
Boxed RT	age, sex, edu[h], and BRT[i]	–0.40	*.01* [f]
Boxed distraction cost	age, sex, edu, and BRT	–0.24	.09
SAAT RT	age, sex, edu, and BRT	–0.01	.92
Spatial Span	age, sex, edu, and BRT	0.34	*.02* [f]

[a]PASAT: Paced Auditory Serial Addition Test.

[b]ACE: Adaptive Cognitive Evaluation.

[c]MS: multiple sclerosis.

[d]RT: reaction time.

[e]N/A: not applicable.

[f]*P* values in italic are significant.

[g]SAAT: Sustained Attention ACE Task.

[h]edu: years of education.

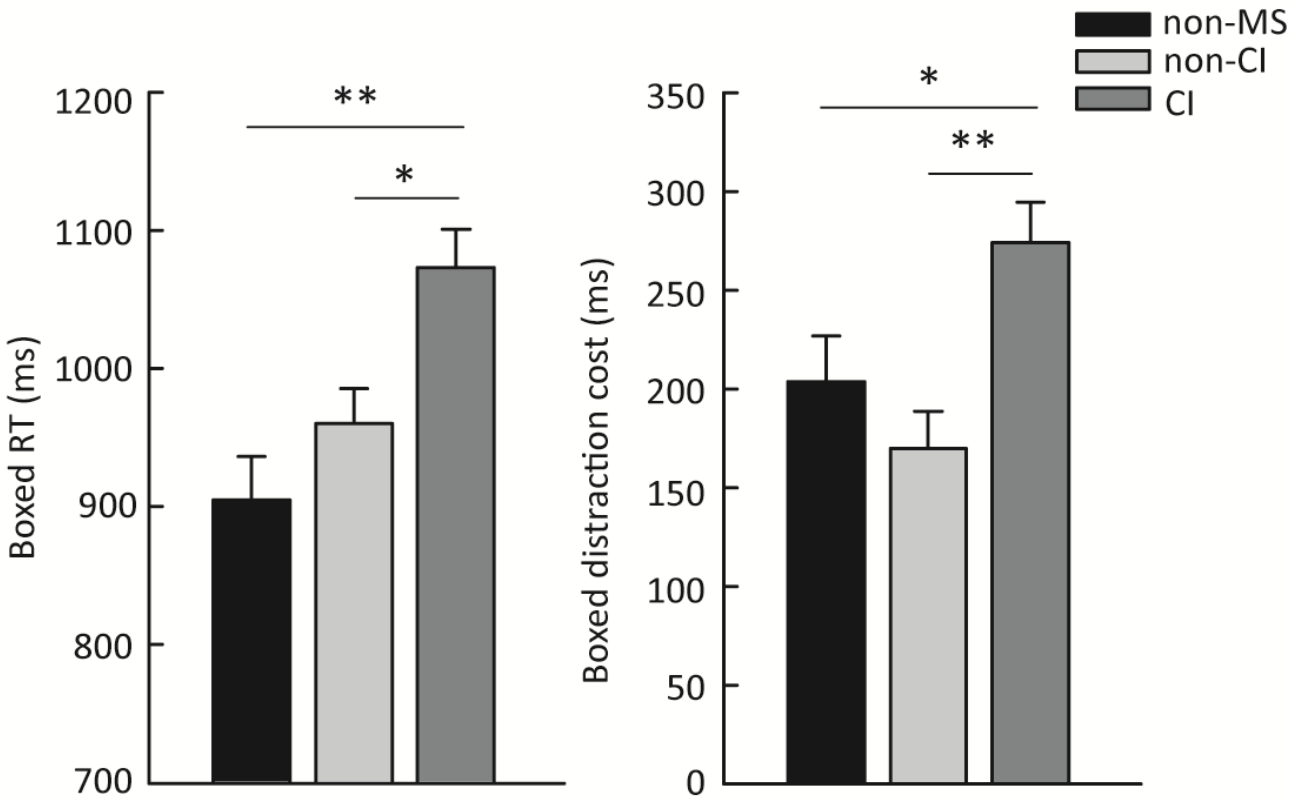
[i]BRT: basic reaction time.

Group Differences in ACE

We then determined whether ACE modules can differentiate participants with MS with CI (SDMT z score<–1) and without CI (SDMT z score≥–1), as well as non-MS participants. To accomplish this, we conducted one-way ANOVA with the participant category as the independent variable for the Boxed RT, Boxed distraction cost, SAAT RT, and Spatial Span. We included age, sex, and years of education as covariates. The BRT was also included as a covariate since there was a significant difference in the BRT among the 3 groups ($F_{(2,71)}=3.96$, $P=.02$), where the non-MS group showed a faster BRT (311.81 [14.50] ms) compared to both CI (362.86 [13.56] ms, $P=.02$) and non-CI (357.27 [12.34] ms, $P=.02$) participants with MS.

Significant group differences in the Boxed RT (F (2,66)=9.73, *P*<.001) and Boxed distraction cost (F (2,66)=7.40, *P*=.001) were revealed. Post hoc comparisons indicated a slower Boxed RT in CI (1072.72 [28.14] ms) compared to non-CI (959.89 [25.48] ms; *P*=.004) participants with MS as well as non-MS participants (904.86 [31.68] ms, *P*<.001). For Boxed distraction cost, CI participants showed a higher attention cost (274.35 [20.47] ms) when compared to non-CI participants (170.12 [18.53] ms, *P*<.001) and non-MS (203.84 [23.04] ms, *P*=.02) participants (Figure 5). Although no statistical differences in Boxed distraction cost were found between the non-CI and non-MS groups (*P*=.20), numerically, the distraction cost in the non-CI group was slightly lower than in the non-MS group. A lower RT cost may be attributed to 2 combinations of task performance: First, the same level of performance for the more challenging task condition but a worse performance in the easier task, and second, the same level of task performance for the easier task and a less performance drop in the more challenging task, indicating better cognitive ability as the performance is not affected much when the task becomes more difficult. Here, the numerical difference in Boxed distraction cost between non-CI and non-MS groups was more likely to be a result of worse performance in the easier task in the non-CI group (ie, slower RT in the 4-item condition) rather than a less performance change in the more challenging task (ie, faster RT in the 12-item condition). To understand this, we investigated RTs in 12-item and 4-item conditions in both groups. As expected, the 2 groups showed the same level of RT in the 12-item condition (non-CI vs non-MS: 1030.22 [24.09] ms vs 1022.23 [30.66] ms, *P*=0.84), but a slower RT was found in the non-CI group in the 4-item condition (non-CI vs non-MS: 854.37 [19.42] ms vs 814.89 [23.87] ms, *P*=0.23). The slower RT in the 4-item condition, to some extent, explained the numerically lower Boxed distraction cost in the non-CI group. No significant difference between the 3 groups was discovered for the SAAT RT (F (2,66)=0.42, *P*=.65) or Spatial Span (F (2,66)=0.62, *P*=.54). The results suggest that the ACE Boxed module can identify group-level differences between CI participants with MS, non-CI participants with MS, and non-MS participants.

Figure 5. Group differences between CI, non-CI participants with MS and non-MS participants. Error bars represent SE.**$P \leqq .001$; *$P \leqq .05$. CI: cognitive impairment; MS: multiple sclerosis; RT: reaction time.



Discussion

Principal Findings

In this study, we aimed to determine whether a digital cognitive assessment battery (ie, ACE) could be used to evaluate cognitive function in adults with MS with and without CI. We found a significant correlational association between ACE metrics and standard cognitive measures. When age, sex, years of education, and the BRT were considered as covariates, only correlations between the SDMT score and the Boxed RT as well as the Boxed distraction cost remained significant. Specifically, the ACE Boxed module, a task measuring visual search performance and attention [38], showed the strongest correlation with the SDMT and could identify group-level differences between adults with MS with and without CI, as well as adults without MS. Altogether, these results provide preliminary evidence that ACE, a tablet-based cognitive assessment battery, provides modules that could potentially serve as an unsupervised cognitive assessment for people with MS.

Correlational links between standard measures and ACE measures were discovered. Specifically, we noted significant correlations between higher SDMT scores and faster Boxed and SAAT RTs, lower Boxed distraction costs, and higher Spatial Span. The significant correlations between SDMT and Boxed measures in all participants (including both non-MS and MS) indicate that this ACE module is a potential cognitive assessment tool, and the results stand alone when only participants with MS are considered. Moreover, an exploratory simple linear regression analysis revealed a moderate R-squared value of 0.33, indicated that 33% of the total variation in the SDMT score can be explained by the Boxed RT. These findings partially support relative consequence theory [30-32], which postulates that a change in the information processing speed is a key deficit underlying cognitive dysfunction in MS [27-30]. However, a clear causal relationship between information processing speed and high-level cognitive performance cannot be concluded with the current results. Of note, when age, sex, years of education, and the BRT were considered as covariates, only correlations between the SDMT score and the Boxed RT and Boxed distraction cost remained significant. Boxed is a visual search task that requires participants to search for a specific target, filter out distractors, and provide a response. To some extent, the task structure and the domains of cognitive function being challenged are similar to those of the SDMT, in which participants are asked to substitute geometric symbols for numbers (make a response) while scanning a response key (search and filter out distractors). The similarity of the cognitive function subserving the 2 tasks may explain the consistent correlation between SDMT scores and Boxed performance. In contrast, the cognitive domains mainly being challenged in SAAT and Spatial Span are attention control and working memory, respectively, which are less similar to what is being challenged during an SDMT test. These different cognitive domains could have been affected differently by factors such as age, sex, and years of education, which could explain the marginally significant correlation revealed between the SDMT and performance in SAAT and Spatial Span when controlling for these factors.

Participants with MS with a higher PASAT score demonstrated a faster Boxed RT and better Spatial Span. PASAT is a test involving information processing, attention, and short-term maintenance and manipulation of information [36]. These associations were expected, as the Boxed module is designed to assess attention and information processing speed, and Spatial Span mainly contains the cognitive component of holding information in mind.

In addition to showing the correlational association between a subset of modules of ACE and standard cognitive measures, we further demonstrated that ACE could differentiate CI in adults with and without MS, as indicated by a significantly slower Boxed RT and the higher attention cost in CI compared to both non-CI participants with MS and non-MS participants. Of note, performance on SAAT and Spatial Span was not significantly different among the 3 groups. Compared to SAAT and Spatial Span, which only challenge 1 or 2 cognitive domains, Boxed is a task that is more complicated and requires more underlying cognitive resources to reach the task goal. Since there is large interindividual variability in the pattern of CI in MS [42,43], it is possible that a complex task requires more executive control and may be a more sensitive tool to capture cognitive dysfunction in participants with MS compared to tasks that challenge only 1 or 2 aspects of cognitive function. These results support our hypotheses that there are correlational links between performance in standard cognitive measures for MS and ACE modules. In addition, as a digital tool in assessing cognitive function in MS, at least 1 ACE module has the capacity to differentiate group-level differences among CI and non-CI MS participants and non-MS participants.

With the advances in digital technology, the assessment and treatment for people with MS have adopted digital platforms [44], which when used in the home can substantially improve accessibility to cognitive remediation programs. Recently, we demonstrated that in-game navigation features of an unsupervised, digital video game–based digital therapeutic could represent a novel and sensitive way to perform cognitive evaluations in MS [34]. The current results further support the use of a digital platform for cognitive assessment in MS. The built-in adaptive algorithms, which modulate task difficulty based on individual performance, reduce interindividual variability, which is usually a concern for cognitive assessments [45,46]. Since ACE is a self-guided digital assessment tool, it has the potential to be used in different settings (eg, at home or in the clinic). Future studies evaluating how ACE performance fluctuates during a day and whether the results would be affected by different testing environments are warranted. Moreover, since each ACE module is designed to challenge specific cognitive components, baseline ACE subtest scores could be useful to inform personalized cognitive training targets. For instance, for a participant with a low-grade score in Spatial Span and a high score in SAAT, the prescribed cognitive training approaches may place great emphasis on working memory rather than attentional control. Studies with a larger sample size or administering ACE as an outcome measure to capture the effect of a cognitive intervention are also needed to provide more information about how the ACE battery can reflect the patient's and the caregiver's real-life experience and to better translate the subtest scores to a meaningful treatment target.

Limitations

Among the limitations of this study, the relatively small sample size and lack of multiple points of data collection at baseline made it difficult to draw a definitive conclusion with respect to the test validity and test-retest reliability of the ACE battery in people with MS. Related to this, given the predominately White participants in the study, particularly those with CI, the potential influence of racial and ethnicity on the results could not be fully excluded. Participants with severe CI were excluded from the study. Although the severity of CI can vary from mild to quite severe in MS patients, it has been reported that the majority (771/1014, 76.03%) of patients experience mild (340/771, 33.7%) to moderate (431/771, 42.7%) cognitive disturbance [47]. Since the application of the ACE program in clinical settings is still at an early stage, we planned to start with patients without severe CI to reduce the heterogeneity of the sample. Future studies with a broader range of CI are needed to investigate how the ACE tool performs when applied to

participants with severe CI. Furthermore, the results of the exploratory simple linear regression analysis should be taken with caution, given the sample size does not have adequate power to provide a rigorous predictive model. Future studies with a larger sample size are warranted for a better predictive model development. Finally, experience with using digital tools may be confounding factors that can impact the results where participants with more experience in using digital tools may have performed better in this study. Future studies investigating digital assessments should control for participants' skills in using digital devices.

Conclusion

In summary, this study suggests that a tablet-based adaptive cognitive battery could be used to perform cognitive assessments in MS. As noted previously [20,34], the high adherence rate indicated that this remote, home-based health care strategy is well accepted by patients with MS, who may have limited access to cognitive assessment or treatment. Since CI poses major limitations to patients with MS, the current findings open up new paths to deploying digital cognitive tests for MS.

Acknowledgments

We would like to thank our research participants. This work was supported by the Doris Duke Charitable Foundation and Akili Interactive.

Conflicts of Interest

RB received research support from the National Multiple Sclerosis Society, the Hilton Foundation, the California Initiative to Advance Precision Medicine, the Sherak Foundation, and Akili Interactive. RB also received personal compensation for consulting from Alexion, Biogen, EMR Serono, Novartis, Pear Therapeutics, Roche Genentech, and Sanofi Genzyme. AG is cofounder, shareholder, board member, and advisor for Akili Interactive Labs, a company that manufactures investigational digital treatments delivered through a video game–like interface.

References

1. Achiron A, Barak Y. Cognitive impairment in probable multiple sclerosis. J Neurol Neurosurg Psychiatry 2003 Apr;74(4):443-446 [FREE Full text] [doi: 10.1136/jnnp.74.4.443] [Medline: 12640060]
2. Rogers JM, Panegyres PK. Cognitive impairment in multiple sclerosis: evidence-based analysis and recommendations. J Clin Neurosci 2007 Oct;14(10):919-927. [doi: 10.1016/j.jocn.2007.02.006] [Medline: 17659875]
3. Hoffmann S, Tittgemeyer M, von Cramon DY. Cognitive impairment in multiple sclerosis. Curr Opin Neurol 2007 Jun;20(3):275-280. [doi: 10.1097/WCO.0b013e32810c8e87] [Medline: 17495620]
4. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. Neurology 1991 May;41(5):685-691. [doi: 10.1212/wnl.41.5.685] [Medline: 2027484]
5. Rae-Grant A, Bennett A, Sanders AE, Phipps M, Cheng E, Bever C. Quality improvement in neurology: multiple sclerosis quality measures; executive summary. Neurology 2015 Nov 24;85(21):1904-1908 [FREE Full text] [doi: 10.1212/WNL.0000000000001965] [Medline: 26333795]
6. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30 [FREE Full text] [doi: 10.1177/1073191116643164] [Medline: 27084835]
7. Anguera JA, Brandes-Aitken AN, Antovich AD, Rolle CE, Desai SS, Marco EJ. A pilot study to determine the feasibility of enhancing cognitive abilities in children with sensory processing dysfunction. PLoS ONE 2017;12(4):e0172616 [FREE Full text] [doi: 10.1371/journal.pone.0172616] [Medline: 28380008]
8. Anguera JA, Brandes-Aitken AN, Rolle CE, Skinner SN, Desai SS, Bower JD, et al. Characterizing cognitive control abilities in children with 16p11.2 deletion using adaptive 'video game' technology: a pilot study. Transl Psychiatry 2016 Sep 20;6(9):e893 [FREE Full text] [doi: 10.1038/tp.2016.178] [Medline: 27648915]
9. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: a proof-of-concept randomized trial. Depress Anxiety 2017 Jun;34(6):508-517 [FREE Full text] [doi: 10.1002/da.22588] [Medline: 28052513]
10. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán PA. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
11. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
12. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [Medline: 27489456]
13. Charvet LE, Yang J, Shaw MT, Sherman K, Haider L, Xu J, et al. Cognitive function in multiple sclerosis improves with telerehabilitation: results from a randomized controlled trial. PLoS ONE 2017 May 11;12(5):e0177177 [FREE Full text] [doi: 10.1371/journal.pone.0177177] [Medline: 28493924]

14. Davis NO, Bower J, Kollins SH. Proof-of-concept study of an at-home, engaging, digital intervention for pediatric ADHD. PLoS ONE 2018;13(1):e0189749 [FREE Full text] [doi: 10.1371/journal.pone.0189749] [Medline: 29324745]
15. Hackett K, Krikorian R, Giovannetti T, Melendez-Cabrero J, Rahman A, Caesar EE, et al. Utility of the NIH Toolbox for assessment of prodromal Alzheimer's disease and dementia. Alzheimers Dement (Amst) 2018;10:764-772 [FREE Full text] [doi: 10.1016/j.dadm.2018.10.002] [Medline: 30505926]
16. Tsoy E, Erlhoff SJ, Goode CA, Dorsman KA, Kanjanapong S, Lindbergh CA, et al. BHA-CS: a novel cognitive composite for Alzheimer's disease and related disorders. Alzheimers Dement (Amst) 2020;12(1):e12042 [FREE Full text] [doi: 10.1002/dad2.12042] [Medline: 32582835]
17. Reijs BLR, Ramakers IHGB, Köhler S, Teunissen CE, Koel-Simmelink M, Nathan PJ, et al. Memory correlates of Alzheimer's disease cerebrospinal fluid markers: a longitudinal cohort study. J Alzheimers Dis 2017;60(3):1119-1128. [doi: 10.3233/JAD-160766] [Medline: 28984585]
18. Kim HS, An YM, Kwon JS, Shin M. A preliminary validity study of the cambridge neuropsychological test automated battery for the assessment of executive function in schizophrenia and bipolar disorder. Psychiatry Investig 2014 Oct;11(4):394-401 [FREE Full text] [doi: 10.4306/pi.2014.11.4.394] [Medline: 25395970]
19. Levaux M, Potvin S, Sepehry AA, Sablier J, Mendrek A, Stip E. Computerized assessment of cognition in schizophrenia: promises and pitfalls of CANTAB. Eur Psychiatry 2007 Mar;22(2):104-115. [doi: 10.1016/j.eurpsy.2006.11.004] [Medline: 17227707]
20. Bove RM, Rush G, Zhao C, Rowles W, Garcha P, Morrissey J, et al. A videogame-based digital therapeutic to improve processing speed in people with multiple sclerosis: a feasibility study. Neurol Ther 2019 Jun;8(1):135-145 [FREE Full text] [doi: 10.1007/s40120-018-0121-0] [Medline: 30506301]
21. Khaligh-Razavi S, Sadeghi M, Khanbagi M, Kalafatis C, Nabavi SM. A self-administered, artificial intelligence (AI) platform for cognitive assessment in multiple sclerosis (MS). BMC Neurol 2020 May 18;20(1):193 [FREE Full text] [doi: 10.1186/s12883-020-01736-x] [Medline: 32423386]
22. Middleton RM, Pearson OR, Ingram G, Craig EM, Rodgers WJ, Downing-Wood H, UK MS Register Research Group. A rapid electronic cognitive assessment measure for multiple sclerosis: validation of cognitive reaction, an electronic version of the Symbol Digit Modalities Test. J Med Internet Res 2020 Sep 23;22(9):e18234 [FREE Full text] [doi: 10.2196/18234] [Medline: 32965240]
23. Rudick RA, Miller D, Bethoux F, Rao SM, Lee J, Stough D, et al. The Multiple Sclerosis Performance Test (MSPT): an iPad-based disability assessment tool. J Vis Exp 2014 Jun 30(88):e51318 [FREE Full text] [doi: 10.3791/51318] [Medline: 25046650]
24. Younger J, O'Laughlin K, Anguera J, Bunge S, Ferrer E, Hoeft F, et al. Development of Executive Function in Middle Childhood: A Large-Scale, In-School, Longitudinal Investigation. 2021 Apr 20. URL: https://psyarxiv.com/xf489/ [accessed 2021-04-20]
25. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R, Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler 2017 Apr 16;23(5):721-733 [FREE Full text] [doi: 10.1177/1352458517690821] [Medline: 28206827]
26. Parmenter BA, Weinstock-Guttman B, Garg N, Munschauer F, Benedict RHB. Screening for cognitive impairment in multiple sclerosis using the Symbol Digit Modalities Test. Mult Scler 2007 Jan;13(1):52-57. [doi: 10.1177/1352458506070750] [Medline: 17294611]
27. Denney DR, Lynch SG. The impact of multiple sclerosis on patients' performance on the Stroop Test: processing speed versus interference. J Int Neuropsychol Soc 2009 May;15(3):451-458. [doi: 10.1017/S1355617709090730] [Medline: 19402931]
28. Drew MA, Starkey NJ, Isler RB. Examining the link between information processing speed and executive functioning in multiple sclerosis. Arch Clin Neuropsychol 2009 Feb;24(1):47-58. [doi: 10.1093/arclin/acp007] [Medline: 19395356]
29. Genova HM, DeLuca J, Chiaravalloti N, Wylie G. The relationship between executive functioning, processing speed, and white matter integrity in multiple sclerosis. J Clin Exp Neuropsychol 2013;35(6):631-641 [FREE Full text] [doi: 10.1080/13803395.2013.806649] [Medline: 23777468]
30. Owens EM, Denney DR, Lynch SG. Difficulties in planning among patients with multiple sclerosis: a relative consequence of deficits in information processing speed. J Int Neuropsychol Soc 2013 May;19(5):613-620. [doi: 10.1017/S1355617713000155] [Medline: 23425634]
31. DeLuca J, Chelune GJ, Tulsky DS, Lengenfelder J, Chiaravalloti ND. Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? J Clin Exp Neuropsychol 2004 Jun;26(4):550-562. [doi: 10.1080/13803390490496641] [Medline: 15512942]
32. Forn C, Belenguer A, Parcet-Ibars MA, Avila C. Information-processing speed is the primary deficit underlying the poor performance of multiple sclerosis patients in the Paced Auditory Serial Addition Test (PASAT). J Clin Exp Neuropsychol 2008 Oct;30(7):789-796. [doi: 10.1080/13803390701779560] [Medline: 18608672]
33. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011 Feb;69(2):292-302 [FREE Full text] [doi: 10.1002/ana.22366] [Medline: 21387374]

34. Bove R, Rowles W, Zhao C, Anderson A, Friedman S, Langdon D, et al. A novel in-home digital treatment to improve processing speed in people with multiple sclerosis: a pilot study. Mult Scler 2021 Apr 25;27(5):778-789. [doi: 10.1177/1352458520930371] [Medline: 32584155]
35. Hsu W, Rowles W, Anguera JA, Zhao C, Anderson A, Alexander A, et al. Application of an adaptive, digital, game-based approach for cognitive assessment in multiple sclerosis: observational study. J Med Internet Res 2021 Jan 27;23(1):e27440 [FREE Full text] [doi: 10.2196/27440] [Medline: 33502997]
36. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills 1977 Apr;44(2):367-373. [doi: 10.2466/pms.1977.44.2.367] [Medline: 866038]
37. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 05;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
38. Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980 Jan;12(1):97-136. [doi: 10.1016/0010-0285(80)90005-5] [Medline: 7351125]
39. Leark RA, Dupuy TR, Greenberg LM, Corman CL, Kindschi C. T.O.V.A. Test of Variables of Attention Professional Manual. Los Alamitos, CA: Universal Attention Disorders; 1999.
40. Berch DB, Krikorian R, Huha EM. The Corsi block-tapping task: methodological and theoretical considerations. Brain Cogn 1998 Dec;38(3):317-338. [doi: 10.1006/brcg.1998.1039] [Medline: 9841789]
41. Kiely KM, Butterworth P, Watson N, Wooden M. The Symbol Digit Modalities Test: normative data from a large nationally representative sample of Australians. Arch Clin Neuropsychol 2014 Dec;29(8):767-775. [doi: 10.1093/arclin/acu055] [Medline: 25352087]
42. Bobholz JA, Rao SM. Cognitive dysfunction in multiple sclerosis: a review of recent developments. Curr Opin Neurol 2003 Jun;16(3):283-288. [doi: 10.1097/01.wco.0000073928.19076.84] [Medline: 12858063]
43. Langdon DW. Cognition in multiple sclerosis. Curr Opin Neurol 2011 Jun;24(3):244-249. [doi: 10.1097/WCO.0b013e328346a43b] [Medline: 21519256]
44. Hoang P, Schoene D, Gandevia S, Smith S, Lord SR. Effects of a home-based step training programme on balance, stepping, cognition and functional performance in people with multiple sclerosis: a randomized controlled trial. Mult Scler 2016 Jan;22(1):94-103. [doi: 10.1177/1352458515579442] [Medline: 25921035]
45. Castellanos FX, Sonuga-Barke EJS, Scheres A, Di Martino A, Hyde C, Walters JR. Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. Biol Psychiatry 2005 Jun 01;57(11):1416-1423 [FREE Full text] [doi: 10.1016/j.biopsych.2004.12.005] [Medline: 15950016]
46. Hetherington CR, Stuss DT, Finlayson MA. Reaction time and variability 5 and 10 years after traumatic brain injury. Brain Inj 1996 Jul;10(7):473-486. [doi: 10.1080/026990596124197] [Medline: 8806008]
47. Harel Y, Kalron A, Menascu S, Magalashvili D, Dolev M, Doniger G, et al. Cognitive function in multiple sclerosis: a long-term look on the bright side. PLoS ONE 2019;14(8):e0221784 [FREE Full text] [doi: 10.1371/journal.pone.0221784] [Medline: 31465498]

Abbreviations

ACE: Adaptive Cognitive Evaluation
BRT: basic reaction time
CI: cognitive impairment
MS: multiple sclerosis
PASAT: Paced Auditory Serial Addition Test
RT: reaction time
SAAT: Sustained Attention ACE Task
SDMT: Symbol Digit Modalities Test
TOVA: Tests of Variables of Attention
UCSF: University of California, San Francisco

Edited by G Eysenbach; submitted 13.11.20; peer-reviewed by A Wright, A Mendez, S Kollins, V Horner; comments to author 13.01.21; revised version received 29.01.21; accepted 16.11.21; published 30.12.21.

Please cite as:
Hsu WY, Rowles W, Anguera JA, Anderson A, Younger JW, Friedman S, Gazzaley A, Bove R
Assessing Cognitive Function in Multiple Sclerosis With Digital Tools: Observational Study
J Med Internet Res 2021;23(12):e25748
URL: https://www.jmir.org/2021/12/e25748
doi:10.2196/25748
PMID:34967751

Original Paper

Evaluating Course Completion, Appropriateness, and Burden in the Understanding Multiple Sclerosis Massive Open Online Course: Cohort Study

Suzi B Claflin[1], PhD; Julie A Campbell[1], PhD; Kathleen Doherty[2], PhD; Maree Farrow[2], PhD; Barnabas Bessing[1], MA; Bruce V Taylor[1], MD

[1]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

[2]Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Australia

Corresponding Author:
Suzi B Claflin, PhD
Menzies Institute for Medical Research
University of Tasmania
17 Liverpool Street
Hobart, 7000
Australia
Phone: 61 0362267728
Email: suzi.claflin@utas.edu.au

Abstract

Background: Massive open online course (MOOC) research is an emerging field; to date, most research in this area has focused on participant engagement.

Objective: The aim of this study is to evaluate both participant engagement and measures of satisfaction, appropriateness, and burden for a MOOC entitled Understanding Multiple Sclerosis (MS) among a cohort of 3518 international course participants.

Methods: We assessed the association of key outcomes with participant education level, MS status, caregiver status, sex, and age using summary statistics, and 2-tailed *t* tests, and chi-square tests.

Results: Of the 3518 study participants, 928 (26.37%) were people living with MS. Among the 2590 participants not living with MS, 862 (33.28%) identified as formal or informal caregivers. Our key findings were as follows: the course completion rate among study participants was 67.17% (2363/3518); the course was well received, with 96.97% (1502/1549) of participants satisfied, with an appropriate pitch and low burden (a mean of 2.2 hours engagement per week); people living with MS were less likely than those not living with MS to complete the course; and people with a recent diagnosis of MS, caregivers, and participants without a university education were more likely to apply the material by course completion.

Conclusions: The Understanding MS MOOC is fit for purpose; it presents information in a way that is readily understood by course participants and is applicable in their lives.

(J Med Internet Res 2021;23(12):e21681) doi:10.2196/21681

KEYWORDS

multiple sclerosis; massive open online course; health promotion; eHealth education; mobile phone

Introduction

Background

Massive open online course (MOOC) research is an emerging field [1,2]. The work done to date has focused on participant engagement, particularly course completion [3], which has presented a challenge for MOOCs because MOOCs have a mean 5% to 15% completion rate [4]. Few studies have evaluated course material appropriateness, participant satisfaction, and reasons for noncompletion. Here, we contribute to this ongoing conversation by evaluating the impact of education level, multiple sclerosis (MS) status, caregiver status, sex, and age on completion, satisfaction, perceived appropriateness, and burden of a MOOC on MS.

MOOCs emerged internationally into the knowledge economy in 2012, where they were heralded as a revolution that would democratize education by offering high-quality courses for free to anyone with access to an internet connection [5]. Since then,

the number of MOOCs has rapidly increased; Class Central, the largest MOOC aggregator website, listed more than 13,000 MOOCs from more than 900 universities in 2019 [6]. However, despite the increased availability of MOOCS, these courses have struggled to reach and retain underserved students in the same numbers as their more privileged peers. As students from more affluent areas are more likely to participate in and complete MOOCs [7,8], MOOCs may exacerbate educational inequalities by offering additional resources to populations that also have access to a range of other educational opportunities.

Health and medicine MOOCs may encounter an additional challenge because they are often developed for use by people living with a health condition and their caregivers to address information asymmetry between the medical profession (as the suppliers) and people with health conditions and their caregivers (as consumers) [9-11]. However, because health status is related to socioeconomic status and education [12,13], people affected by a health condition may be less likely to enroll and complete a MOOC than those who are unaffected. Fortunately, previous studies suggest that these challenges can be addressed successfully. The Wicking Dementia Research and Education Centre (WDREC) has developed a MOOC on dementia that demonstrably improves knowledge of dementia in participants with a wide range of educational attainment [14,15], indicating that appropriately designed MOOCs can overcome some of these barriers.

Using the WDREC MOOCs as a successful model of knowledge dissemination, we have developed a free 6-week MOOC about MS [16] to increase awareness and understanding of MS in the MS community and interested laypeople. MS is a chronic autoimmune disorder where the immune system attacks and damages the central nervous system [17]. MS-related symptoms, such as mobility impairment and fatigue, may make it difficult for people living with MS to access traditional educational offerings [18]. After a year of development in collaboration with the MS community (eg, people with MS, carers, service providers, health care providers, and researchers), the Understanding MS MOOC was released in 2019 and had 2 open enrollments in that year. It was well received by participants, ranking first among the >2400 MOOCs released in 2019 based on participant reviews [6,19].

Objective

In this study, we assessed the impact of the course on information asymmetry in the MS community. In health care (particularly in health services and health economics), information asymmetry (or asymmetry of information) relates to the difference in the information known by the consumer (eg, the patient or a member of the public) and that known by the producer or supplier, a health care professional [20]. In the information age, the gap creating information asymmetry could close if consumers can access appropriately pitched, validated, and targeted information sources [21]. Therefore, to assess the potential impact of the course on information asymmetry, we explored the overall course completion rate, participant satisfaction, perceived appropriateness and burden, and the association between these outcomes and demographic and health factors.

Methods

Overview

The data for this study were collected during the 2 enrollments of the Understanding MS MOOC administered in 2019. The course is free and available in English internationally on any internet-connected device (eg, computer or smartphone; [22]). Course content is presented in videos (transcripts are available for all videos), text, images, and animations. The content is presented in 6 modules over 6 weeks, and course participants can access the material for a total of 8 weeks. Each module contains at least 1 optional activity and discussion prompt. At the end of each module is a summary of the module content and a 10-question multiple-choice quiz. Participants can take the quiz as many times as they like but must achieve a score of 70% or higher to move on to the next module. The course covers topics ranging from the underlying pathology of MS to its impact on everyday life and includes both academic content and lived experience videos from a range of MS community members (for a more detailed description, refer to the study by Claflin et al [16]).

An optional feedback survey was accessible in the completion section during the 2- to 3-week period that the section was open before course closure. Therefore, the survey was only available to the participants who completed the course. We chose to place the feedback survey in the completion section to ensure that all survey respondents had completed the full intervention. An analysis of reasons for noncompletion is underway in a separate study. The feedback survey was adapted from a similar tool used to assess a WDREC MOOC about dementia [14] and queried participants' overall satisfaction with the course and various aspects of the course. With a few exceptions, the questions in this survey were presented on a 5-point Likert scale, ranging from very dissatisfied to very satisfied, or strongly disagree to strongly agree and example survey questions are available in the study by Claflin et al [16].

Small changes to the web-based content were made between the 2 enrollments based on feedback from the first enrollment. We added 3 short videos (<3 minutes each): 1 on exercise physiology, 1 on physical therapy, and 1 on comorbidities. We added a couple of paragraphs of text about disease-modifying therapies and more clearly identified the activities in each module. We also added 2 small interactive features to help participants navigate through a series of short videos on symptoms and risk factors.

The course was advertised widely through social media, particularly through Facebook ads. Advertising targeted anglophone countries. Information about the course was also disseminated through the Menzies Institute network, as well as that of our project partners, Multiple Sclerosis Limited and WDREC, and other related organizations.

Participants in this study gave informed consent for their course-collected data, including their course feedback survey, to be used for research purposes in the introduction or orientation section of the course before they had access to any course content. This study was approved by the University of

Tasmania Social Science Human Research Ethics Committee (H0017892).

Demographic and Health Status Characteristics

This study evaluated 3 primary predictor variables for course completion and course satisfaction: MS status, caregiver status, and education level. These variables were of primary importance, as education has been shown to affect course completion in many MOOCs, and as a course intended for the MS community and interested laypeople, the course can only be considered fit for purpose if it is appropriate for people with MS and their caregivers.

Participants self-reported demographic and health status characteristics during course enrollment and in the feedback survey. This includes self-identification with various roles in the MS community. We categorized all participants into 2 MS status groups, as people with MS or those not living with MS, based on this information. We categorized people not living with MS into 2 caregiver status groups: not caregivers and caregivers, defined as anyone who identified as either a family or friend of a person with MS or a caregiver, thereby incorporating both formal and informal caregivers into a single group.

Similarly, participants self-reported their education level as grade 12 or below, occupational certificate or diploma, undergraduate degree, or postgraduate degree. We then categorized all participants into 2 education-level groups: no university education (grade 12 or below and occupational certificate or diploma) and university education (undergraduate or postgraduate degree), following the methodology of Goldberg et al [14].

Our secondary predictors were MS disease duration, sex, and age, which were self-reported during course enrollment. We calculated age from self-reported year of birth and calculated MS disease duration from self-reported year of diagnosis.

Outcome Measures

Completion

We evaluated participant completion using course-collected data and compared the completion rate with the average for MOOCs, which is 5% to 15% [4]. We determined the course completion and final course module using quiz attempts. Any attempt to complete a quiz (whether or not the score was sufficient to move on to the next module) was considered an indication that the participant had completed the module. All module 6 (final module) quiz attempts were considered an indication of course completion. We assessed the association between completion rate and demographic and health status characteristics (MS status, caregiver status, education level, sex, age, and disease duration).

Satisfaction, Perceived Appropriateness, and Burden

We evaluated satisfaction, perceived appropriateness, and burden among course completers using data from the course feedback survey.

Satisfaction was determined using 2 questions. The first was a 5-point Likert scale that asked about overall satisfaction with the course. We categorized responses into 2 groups: satisfied (satisfied or very satisfied) or not satisfied (neutral, dissatisfied, or very dissatisfied). The second was a 6-point Likert scale that asked about the overall course quality. We categorized responses into 2 groups: average or below (average, poor, very poor), or above average (good, very good, and excellent).

We evaluated the appropriateness and burden of the course with questions querying (1) self-reported agreement that the participant could understand the content, (2) that the language was too technical, (3) that there was too much or too little material, (4) that the course improved their understanding, (5) that the material could improve care or quality of life for people with MS, (6) that they would recommend the course, and (7) that they had already applied course material in their lives. Responses were categorized into 2 groups: agree (agree or strongly agree) and disagree (neutral, disagree, or strongly disagree). We also assessed the burden by comparing the self-reported average time spent on a single course module between groups.

Analysis

We cleaned the data set by removing any staff accounts and removing the second attempts of any participant who took part in both enrollments. During data cleaning, we designated ages (based on self-reported year of birth) of <10 years or >95 years as no data, as these values were deemed implausible. Similarly, we excluded impossible or uninterpretable years of diagnosis (eg, 1 or 1900).

As this data set was overpowered, there were many statistically significant differences that were not of interest because they were not reflective of materially significant differences between groups. To account for this, we set a threshold of material significance for comparisons between categorical variables, which required a 5% difference between groups. We report the results of these materially significant differences (all of which are statistically significant). As age and disease duration were continuous variables, we evaluated their effects on all outcomes of interest. To determine whether the enrollments could be evaluated together, we compared the outcomes of interest to assess if there were any materially significant differences (>5%).

We assessed the demographics of study participants using the sample size and percentage of the cohort for categorical variables and mean and SD for continuous variables. We assessed the association between the predictor variables on the responses of interest and the relationships between the predictor variables using 2-tailed chi-square and *t* tests. As disease duration was not normally distributed, we evaluated its association with the outcomes of interest using the Mann-Whitney rank-sum tests of equal medians. We used Pearson correlation to evaluate the association between the average time taken to complete a module and participant age and disease duration. In all analyses, statistical significance was set at $P<.05$. All analyses were conducted using STATA (version 16.0, StataCorp).

Results

Participant Characteristics

In total, 8324 people enrolled in the first 2 enrollments of the Understanding MS MOOC, 3912 (46.99%) of whom completed the course. After removing 52 second attempts, 3518 unique participants across the 2 course enrollments gave permission for their data to be used in research; 1549 consenting course completers also completed a feedback survey (Figure 1).

Participant characteristics are presented in Table 1. Differences between enrollments were <5% for the outcomes of interest. Therefore, data from both enrollments were assessed together.

The majority of the participants were women and had an undergraduate degree or higher education level (course data in Table 1). Nearly two-thirds of the participants resided in Australia, with other large anglophone countries with high MS prevalence comprising the other most well-represented nations in the sample (eg, Canada and New Zealand). Nearly a third of the participants were people with MS. Of the 2096 course participants not living with MS, 862 (41.12%) identified as formal or informal (family or friends of people with MS) caregivers.

Figure 1. Inclusion flowchart.

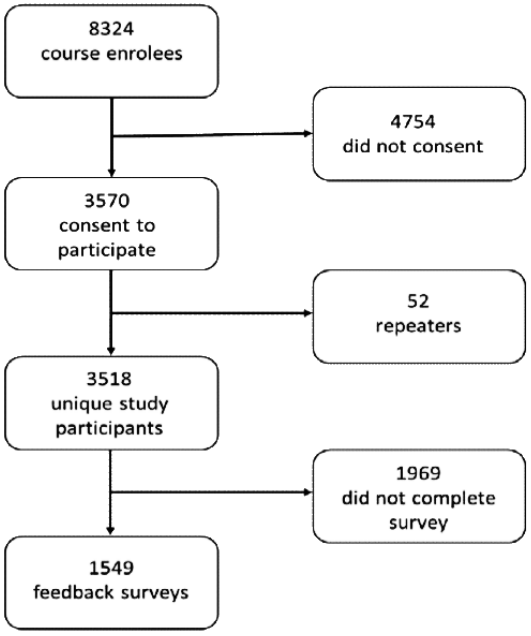


Table 1. Characteristics of study participants who provided course-collected data and course completers who supplied feedback surveys. Please note that participants could select multiple MS[a] community roles[b].

Characteristic	Course-collected data	Feedback survey
Gender, n (%)	3492 (100)	1412 (100)
Male	552 (15.81)	227 (16.08)
Female	2940 (84.19)	1185 (83.92)
Education level, n (%)	2876 (100)	1549 (100)
No university education	1169 (40.65)	595 (38.41)
University education	1707 (59.35)	954 (61.59)
MS community roles[c], n (%)	3024 (100)	1549 (100)
Person with MS	928 (30.69)	437 (28.21)
Family member or friend	664 (21.96)	382 (24.66)
Carer	352 (11.64)	139 (8.97)
Service provider	360 (11.9)	92 (5.94)
Allied health	815 (26.95)	365 (23.56)
General practitioner	67 (2.22)	41 (2.65)
Neurologist	62 (2.05)	18 (1.16)
Advocate	67 (2.22)	44 (2.84)
Researcher	123 (4.07)	44 (2.84)
Other or no MS community role	374 (12.37)	321 (20.72)
Country of residence, n (%)	3509 (100)	1417 (100)
Australia	2180 (62.13)	907 (64.01)
Canada	100 (2.85)	38 (2.68)
United Kingdom	255 (7.27)	101 (7.13)
Ireland	133 (3.79)	44 (3.11)
New Zealand	277 (7.89)	127 (8.72)
United States	106 (3.02)	35 (2.47)
South Africa	51 (1.45)	27 (1.91)
Other	407 (11.6)	138 (9.74)
Final section completed, n (%)	3518 (100)	—[d]
<Module 1	620 (17.62)	—
Module 1	251 (7.13)	—
Module 2	90 (2.59)	—
Module 3	74 (2.1)	—
Module 4	72 (2.05)	—
Module 5	48 (1.36)	—
Module 6	2363 (67.17)	—
Age (years), mean (SD)	44.38 (13.34)[e]	46.78 (13.10)[f]
Disease duration (years), median (SD)	4 (10)	5 (10)[g]

[a]MS: multiple sclerosis.

[b]Among people with multiple sclerosis, approximately half of the participants who provided course-collected data and those who provided feedback data had a disease duration of 4 years or less. Consequently, the distribution was highly skewed toward 0 years (diagnosis in 2019; Multimedia Appendix 1).

[c]Multiple selections possible.

[d]Not available.
[e]N=3292.
[f]N=1330.
[g]N=401.

Completion

Of the 3518 course participants who gave permission for their course-collected data to be used in research, 2363 (67.17%) completed the course. There were significant differences in the completion rate between MS status groups (χ^2_1=36.8; *P*<.001; Table 2). People with MS were less likely to complete the course than those not living with MS (539/928, 58.08% compared with 1455/2096, 69.42%). People with MS completed an average of 3.9 modules while those not living with MS completed an average of 4.5. This association was consistent across all course modules, with people with MS less likely to complete module 1 (χ^2_1=26.7; *P*<.001), module 2 (χ^2_1=22.2; *P*<.001), module 3 (χ^2_1=26.03; *P*<.001), module 4 (χ^2_1=28.8; *P*<.001), and module 5 (χ^2_1=27.9; *P*<.001).

Table 2. The percentage of participants completing the course, satisfied with the course, or agreeing with various statements about the course in the course feedback survey in different participant groups, and the absolute difference between groups. Italicized values indicate materially significant (>5%) differences between groups.

Participant groups	Course completion (%)	Course feedback survey data (%)										
		Satisfied	Above average quality	Improved understanding	Could understand	Language too technical	Too much material	Not enough material	Can improve care	Can improve quality of life	Would recommend	Already applied
University education, n (%)												
No	744 (63.64)	582 (97.82)	583 (98.31)	562 (96.1)	572 (97.28)	47 (8.06)	28 (4.79)	69 (11.82)	551 (94.03)	528 (90.26)	572 (98.28)	391 (66.95)
Yes	1162 (68.07)	920 (96.44)	936 (98.42)	874 (93.6)	919 (98.08)	65 (6.96)	48 (5.17)	152 (16.34)	856 (92.04)	829 (89.72)	899 (96.98)	554 (61.15)
|x-y|[a](%)	4.43	1.38	0.11	2.49	0.80	1.10	0.37	4.53	1.98	0.54	1.30	*5.80*[b]
MS[c] status, n (%)												
People with MS	539 (58.08)	420 (96.11)	421 (97)	387 (89.79)	418 (96.76)	32 (7.44)	22 (5.14)	67 (15.55)	383 (89.07)	369 (87.03)	413 (96.72)	281 (66.59)
People not living with MS	1455 (69.42)	1082 (97.3)	1098 (98.92)	1049 (96.42)	1073 (98.17)	80 (7.36)	54 (4.98)	154 (14.22)	1024 (94.29)	988 (91.06)	1038 (97.78)	664 (62.17)
|x-y| (%)	*11.34*[b]	1.19	1.91	*6.62*[b]	1.41	0.08	0.16	1.33	*5.22*[b]	4.03	1.06	4.42
Caregiver status, n (%)												
No	877 (71.07)	672 (97.39)	678 (98.55)	646 (95.99)	666 (98.09)	49 (7.25)	33 (4.9)	101 (15.05)	630 (93.75)	610 (90.77)	654 (97.32)	388 (58.7)
Yes	578 (67.05)	410 (97.16)	420 (99.53)	403 (97.11)	407 (98.31)	31 (7.54)	21 (5.12)	53 (12.86)	394 (95.17)	378 (91.53)	404 (98.54)	276 (67.81)
|x-y| (%)	4.02	0.23	0.98	1.12	0.22	0.29	0.21	2.19	1.42	0.75	1.22	*9.11*[b]
Sex, n (%)												
Female	1985 (67.52)	1149 (96.96)	1166 (98.56)	1098 (94.33)	1150 (98.29)	73 (6.28)	48 (4.13)	166 (14.27)	1086 (93.38)	1043 (89.99)	1127 (97.41)	728 (63.69)
Male	366 (66.30)	221 (97.36)	221 (98.22)	210 (95.45)	210 (95.02)	26 (11.76)	21 (9.63)	39 (17.73)	200 (91.74)	196 (90.74)	214 (97.72)	139 (64.65)
|x-y| (%)	1.21	0.39	0.34	1.12	3.27	*5.48*[b]	*5.51*[b]	3.45	1.64	0.75	0.36	0.96

[a]|x-y|: absolute difference between groups.
[b]Materially significant difference level was set at >5%.
[c]MS: multiple sclerosis.

To further explore this, we evaluated the completion rate among participants who completed module 1 and found that the difference between those living with MS and those not living with MS shrank to about 6% (539/707, 76.24% compared with 1455/1762, 82.58%). This difference was maintained among those who completed module 2. Among module 3 completers, the difference between people with MS and those not living with MS dropped below the threshold for material significance and continued to decline in the remaining module completion groups. There were no materially significant differences in completion based on caregiver status, sex, or education level, but age was significantly associated with completion. Course completers were more likely to be older than noncompleters (Table 3). However, the effect size was not large; the mean age of completers was 45 years compared with 42 years for noncompleters. Similarly, among people with MS, participants with more recent diagnoses were less likely to complete the course than those who had been living with MS for longer periods (Table 3). However, the effect size was small (median disease duration of 1 year compared with 2 years).

Table 3. Results of *t* tests, Mann-Whitney rank-sum tests of equal medians, and Pearson correlations evaluating the association between age and disease duration, and all outcome variables.

Participant groups	Course completion	Course feedback survey data											Average hours to complete 1 module[a]	
		Satisfied	Above average quality	Improved understanding	Could understand	Language too technical	Too much material	Not enough material	Can improve care	Can improve quality of life	Would recommend	Already applied	Coefficient	*P* value
Age													0.06	*.03*[b]
t test	−6.26 (3290)	−0.052 (1328)	−0.32 (1325)	−1.55 (1304)	−0.52 (1308)	4.55 (1301)	4.17 (1299)	2.38 (1301)	0.16 (1298)	0.42 (1292)	−1.41 (1293)	1.27 (1277)		
P value	*<.001*[b]	.60	.75	.12	.60	*<.001*[b]	*<.001*[b]	*.02*[b]	.87	.67	.16	—[c]		
Multiple sclerosis disease duration													−0.023	.65
z	−2.154	−0.304	0.551	2.349	0.231	−0.083	0.694	2.391	2.027	0.865	0.158	3.078		
P value	*.03*[b]	.77	.59	*.02*[b]	.82	.93	.49	*.02*[b]	*.04*[b]	.39	.88	*.002*[b]		

[a]Estimates from Pearson correlation.

[b]Indicate *P* values <.05.

[c]Not available.

Satisfaction, Perceived Appropriateness, and Burden

Overall, course completers were satisfied with the course, with 96.97% (1502/1549) of those completing the feedback survey reporting that they were satisfied or very satisfied (Multimedia Appendix 2). They also rated the quality of the course highly, with 98.38% (1519/1544) rating it above average (good, very good, or excellent). The pitch of the course appears appropriate, with nearly all participants agreeing that they could understand the information, that the course improved their understanding, and that they would recommend the course.

Participants also found the material helpful, with 63.42% (945/1490) reporting that they had applied information from the course at course completion, and nearly all) agreed that the information could improve care (1407/1516, 92.81%) or quality of life (1357/1509, 89.93%) for people with MS. In addition, the burden was low (average of 2.2 hours to complete a module). Only 5.02% (76/1513) agreed that there was too much material, whereas 14.6% (221/1514) agreed that there was too little material.

There were few materially significant differences in the responses of the demographic and health status groups (Table 2). People with MS were less likely to report improved understanding because of the course material (χ^2_1=26.2; *P*<.001) and were less likely to agree that the course material could improve care (χ^2_1=12.6; *P*<.001). Among people not living with MS, caregivers were more likely to report applying the course material by course completion than noncaregivers (χ^2_1=60.0; *P*<.001).

University education was also associated with applying the course material; participants with a university education were less likely to report applying the course material at course completion than those without one (χ^2_1=5.2; *P*=.02). Sex was significantly associated with agreement that the language in the course was too technical (χ^2_1=8.4; *P*=.004) and that there was too much material (χ^2_1=11.7; *P*<.001). Male participants were more likely to agree with these statements than female participants. However, there was no difference in the average time spent per module between males and females.

Age was associated with several outcomes of interest (Table 3). Participants who agreed that there was too much material in the course were more likely to be older (mean age of 47 years compared with 40 years). Correspondingly, those who agreed that there was not enough material were more likely to be

younger. Participants who agreed that the language was too technical were also more likely to be older. However, the effect sizes for the latter 2 associations were small, with differences in mean age between 2.5 and 3.5 in the 2 groups. Similarly, increasing age was associated with a greater average number of hours taken to complete a module, but the effect size was small (coefficient=0.02; Table 3).

Among people with MS, disease duration was also associated with several outcomes of interest (Table 3). People with MS who agreed that the course had improved their understanding were more likely to have a shorter disease duration than those who did not (median disease duration of 1 year compared with 4 years). A total of 93.5% (172/184) of participants with disease durations of ≤4 years reported improved understanding, compared with 86.2% (181/210) of those with disease durations of >4 years. Similarly, participants who reported that they had applied information from the course by course completion were more likely to be recently diagnosed (median of 1 year compared with 3 years). A total of 73.9% (136/184) participants with a disease duration of ≤4 years reported applying the course material, compared with 58.9% (46/78) of the participants with a disease duration >4 years.

Participants with more recent diagnoses were also more likely to report that there was not enough material in the course than those with older diagnoses (median disease duration of 1 year compared with 2 years) and that the content of the course could improve care for people with MS (median of 1 year compared with 2 years), although the effect sizes of these comparisons were small (1 year).

Associations Between Demographic and Health Status Characteristics

Among course completers, education level was associated with MS status, caregiver status, and sex (Table 4). People with MS were less likely than participants not living with MS (230/437, 52.63% compared with 724/1112, 65.11%) to have a university education. Among those not living with MS, caregivers were less likely than noncaregivers (215/422, 50.95% compared with 509/690, 73.77%) to have a university education. Males were more likely than females to have university education (158/227, 69.6%) compared with 61.01% (723/1185).

Age was significantly associated with caregiver status and education level (Table 4). Participants who were caregivers were more likely to be older than noncaregivers (mean age 50 years compared with 44 years). Similarly, participants without a university education were more likely to be older than those with a university education (mean of 50 years compared with 45 years). Among people with MS, MS disease duration was not associated with sex or education level but was strongly associated with age (Table 4).

Table 4. Results of chi-square and *t* tests evaluating the associations between demographic and health status groups.

Participant groups	University education		MS[a] status		Caregiver status		Sex	Age, coefficient (*P* value)[b]
	No	Yes	People with MS	People not living with MS	No	Yes		
MS status[c], n (%)								
People with MS	207 (47.37)	230 (52.63)	—[d]	—	—	—	—	—
People not living with MS	388 (34.89)	724 (65.11)	—	—	—	—	—	—
Caregiver status[e], n (%)								
No	181 (26.23)	509 (73.77)	—	—	—	—	—	—
Yes	207 (49.05)	215 (50.95)	—	—	—	—	—	—
Sex[f], n (%)								
Female	462 (38.99)	723 (61.01)	350 (29.54)	835 (70.46)	513 (61.44)	322 (38.56)	—	—
Male	69 (30.40)	158 (69.60)	69 (30.40)	158 (69.60)	107 (67.72)	51 (32.28)	—	—
Age								
t test	6.05 (1328)		−1.80 (1328)	—	−7.15 (934)	—	0.93 (1323)	—
P value	*<.001*[g]		.07	—	*<.001*[g]	—	.35	—
Disease duration[h]								
z	0.533		—	—	—	—	−0.556	0.461
P value	.59		—	—	—	—	.58	*<.001*[g]

[a]MS: multiple sclerosis.

[b]Estimates from Pearson correlation.

[c]χ^2_1=20.6; *P*<.001

[d]Not available.

[e]χ^2_1=60.03; *P*<.001

[f]University education: χ^2_1=5.99; *P*=.01; multiple sclerosis status: χ^2_1=0.07; *P*=.80; caregiver status: χ^2_1=2.2; *P*=.14

[g]Indicate *P* values <.05.

[h]Among people with multiple sclerosis.

Discussion

Principal Findings

To our knowledge, the Understanding MS web-based course is the largest MS-related web-based course in the world. To date, more than 13,000 people from 128 countries have enrolled in the course, and it was ranked first among the >2400 courses released in 2019 based on participant reviews. Correspondingly, we found that overall, the Understanding MS MOOC had a completion rate that was more than 3 times higher than the average for MOOCs and very high participant satisfaction. However, there were materially significant differences in participant experience among the different participant groups. People with MS were less likely than those not living with MS to complete the course. Although 63.42% (945/1490) of all course completers reported applying the course material, caregivers and those without a university education were more likely to apply it. Overall, the Understanding MS MOOC is fit for purpose, with an appropriate pitch and burden level, and presents information that is relevant to participants' lives. By disseminating relevant content directly to information consumers (people with MS, caregivers, and those without a university education), the course addresses information asymmetry in the MS community.

Completion

The first 2 open enrollments of the Understanding MS MOOC had an average completion rate of 47% (data not presented here). However, among study participants (the subset of all course participants who consented to take part in this research), there was a 67.17% (2363/3518) completion rate. This is 3-9 times higher than the average for all MOOCs, which fluctuates between 5% and 15% [4]. Course completion was about 11% higher among those not living with MS (539/928, 58.08%) than among those with MS (1455/2096, 69.42%), driven by noncompletion early in the course, particularly in module 1. This may be because of the additional challenges faced by people with MS that may interfere with their ability to complete the course, including complications arising from MS-related symptoms such as fatigue and cognitive impairment.

In addition, course completion was about 4% higher among participants with a university education (1162/1707, 68.07%) than among those without (744/1169, 63.64%). Although this difference is not materially significant, it is larger than other similar courses, such as the Understanding Dementia MOOC developed by WDREC [4], who observed a difference of 0.44% between groups. This discrepancy may be because of the underlying differences in the course participants. The Understanding Dementia MOOC is intended primarily for dementia carers rather than those living with dementia, whereas the Understanding MS MOOC is aimed at a broad audience, including people with the condition; 30.69% (928/3024) of this sample comprised people with MS. People with MS were both less likely to have completed university and less likely to complete the Understanding MS MOOC. The difference in completion between education levels may reflect the difference in completion rate associated with MS status. The data support the possibility that education level and health status interactively affect completion. Among the study participants, people with MS without a university education had the lowest completion rate (250/450, 55.56%) of any MS status or education level group. Conversely, people not living with MS who had a university education had the highest completion rate (904/1278, 70.74%).

Satisfaction, Perceived Appropriateness, and Burden

Among course completers, satisfaction and perceived appropriateness were high in all demographic and health status groups, with >98% satisfied and ≥95% agreeing that they could understand the course material. This agrees with previous work on health and medicine MOOCs, which found >80% participant satisfaction among allied health professionals (Harvey et al 2014 [13]) and members of the community (Tieman et al 2018 [23]). The course also presents a low burden for participants, with participants reporting that the material took an average of 2.2 hours per week to complete. This is far lower than the average 4.2 hours per week required by health and medicine MOOCs [24].

Almost two-thirds of the course completers reported applying course material by completion. However, there were significant differences in the application of course materials between the participant groups. People with MS who were newly diagnosed, caregivers, and those without a university education were more likely to report that they had applied the course material. Newly diagnosed people with MS, who were also more likely to report that the course improved their understanding, were well positioned to apply the course material immediately. Among people not living with MS, caregivers may be better positioned to apply the material immediately. Again, the association between caregiver status and education level, with caregivers less likely to have a university education than noncaregivers, may in part drive the observed association between education level and application of course material. The data support this; caregivers without a university education were the most likely to report applying information by completion (140/201, 69.65%) of any caregiver or education level group. Noncaregivers with university education were least likely to report applying it (279/483, 57.76%). However, this result also agrees with the large body of work demonstrating that higher education levels are associated with higher health literacy and better health outcomes [13,25]. Participants with lower education levels may have lower baseline health literacy and MS-related knowledge, and therefore, learn more from the course.

Knowledge Dissemination to Address Information Asymmetry

Health information is a valuable commodity. High-value health care relies on effective information exchange [11], and better information dissemination is needed to close the gap between health information providers and information consumers [10]. This study demonstrates that participation in the Understanding MS MOOC helps to address information asymmetry among course completers. By course completion, participants successfully translated information by applying it to their lives. This is particularly clear among newly diagnosed people with MS (disease duration of 0-4 years) and caregivers, who are the most likely to apply the course material by completion (136/182, 74.73% and 276/407, 67.81%) reported applying course material by completion, respectively). Recent research suggests that services intended for caregivers need to be sensitive to the fluctuating demands placed upon caregivers and be flexible in their support [23]. This study shows that the Understanding MS MOOC accommodates the needs of health information consumers, such as newly diagnosed and caregivers, and can help to address information asymmetry in the MS community.

Strengths, Limitations, and Future Directions

The main strength of this study is the large and diverse course evaluation cohort. The study participants comprised an international cohort of MS community members and interested laypeople with a range of educational attainment. This study had 2 main limitations. First, the analysis grouped formal and informal carers. These groups may have different needs and characteristics that we were unable to parse in this study. Second, the course evaluation survey was only presented to course completers, making the group vulnerable to selection bias. Although we cannot control for this bias, we have presented our results accordingly. Future research should explore the impact of MS status on the reasons for noncompletion.

Conclusions

The Understanding MS MOOC is an accessible health education intervention with a pitch and burden that is appropriate for course participants. It presents information relevant to the lives of the participants and can be immediately applied. Because a large proportion of course participants identify with MS community roles that are traditional consumers of information, the results of this study suggest that the Understanding MS courses can help to address information asymmetry.

Acknowledgments

The authors would like to thank the Understanding Multiple Sclerosis massive open online course participants who participated in this study and Multiple Sclerosis Limited for supporting the course.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Disease duration (years) among people with multiple sclerosis who participated in this study (n=825).
[PNG File , 26 KB - jmir_v23i12e21681_app1.png]

Multimedia Appendix 2
Percent agreement among course feedback survey participants in response to various statements about the course (exact percentage provided in data labels).
[PNG File , 35 KB - jmir_v23i12e21681_app2.png]

References

1. Al-Rahmi W, Aldraiweesh A, Yahaya N, Bin Kamin Y, Zeki AM. Massive Open Online Courses (MOOCs): data on higher education. Data Brief 2018 Dec 1;22:118-125 [FREE Full text] [doi: 10.1016/j.dib.2018.11.139] [Medline: 30581914]
2. Appleton D, Robertson N, Mitchell L, Lesley R. Our disease: a qualitative meta-synthesis of the experiences of spousal/partner caregivers of people with multiple sclerosis. Scand J Caring Sci 2018 Dec;32(4):1262-1278. [doi: 10.1111/scs.12601] [Medline: 30144143]
3. Claflin S, Gates R, Maher M, Taylor BV. Building a successful massive open online course about multiple sclerosis: a process description. J Med Internet Res 2020 Jul 29;22(7):e16687 [FREE Full text] [doi: 10.2196/16687] [Medline: 32723726]
4. Culyer A. The Dictionary of Health Economics, Third Edition. Cheltenham, United Kingdom: Edward Elgar Publishing; 2014.
5. Daniel J. Massive Open Online Courses: what will be their legacy? FEMS Microbiol Lett 2016 Apr;363(8):fnw055. [doi: 10.1093/femsle/fnw055] [Medline: 26940290]
6. Dobson R, Giovannoni G. Multiple sclerosis - a review. Eur J Neurol 2019 Jan;26(1):27-40. [doi: 10.1111/ene.13819] [Medline: 30300457]
7. Dubay LC, Lebrun LA. Health, behavior, and health care disparities: disentangling the effects of income and race in the United States. Int J Health Serv 2012;42(4):607-625. [doi: 10.2190/HS.42.4.c] [Medline: 23367796]
8. Ebben M, Murphy J. Unpacking MOOC scholarly discourse: a review of nascent MOOC scholarship. Learn Media Technol 2014;39(3):328-345. [doi: 10.1080/17439884.2013.878352]
9. Eccleston C, Doherty K, Bindoff A, Robinson A, Vickers J, McInerney F. Building dementia knowledge globally through the Understanding Dementia Massive Open Online Course (MOOC). NPJ Sci Learn 2019 Apr 10;4:3 [FREE Full text] [doi: 10.1038/s41539-019-0042-4] [Medline: 30993003]
10. Goldberg LR, Bell E, King C, O'Mara C, McInerney F, Robinson A, et al. Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. BMC Med Educ 2015 Mar 26;15:60 [FREE Full text] [doi: 10.1186/s12909-015-0344-z] [Medline: 25889490]
11. Hahn RA, Truman BI. Education improves public health and promotes health equity. Int J Health Serv 2015;45(4):657-678 [FREE Full text] [doi: 10.1177/0020731415585986] [Medline: 25995305]
12. Hansen JD, Reich J. Democratizing education? Examining access and usage patterns in massive open online courses. Science 2015 Dec 04;350(6265):1245-1248. [doi: 10.1126/science.aab3782] [Medline: 26785488]
13. Harvey LA, Glinsky JV, Lowe R, Lowe T. A massive open online course for teaching physiotherapy students and physiotherapists about spinal cord injuries. Spinal Cord 2014 Dec;52(12):911-918. [doi: 10.1038/sc.2014.174] [Medline: 25330938]
14. MOOC completion rates: the data. katyjordan. URL: http://www.katyjordan.com/MOOCproject.html [accessed 2020-02-27]
15. Labelle R, Stoddart G, Rice T. A re-examination of the meaning and importance of supplier-induced demand. J Health Econ 1994 Oct;13(3):347-368. [doi: 10.1016/0167-6296(94)90036-1] [Medline: 10138860]
16. Liyanagunawardena T, Williams SA. Massive open online courses on health and medicine: review. J Med Internet Res 2014 Aug 14;16(8):e191 [FREE Full text] [doi: 10.2196/jmir.3439] [Medline: 25123952]
17. Pearson JF, Alla S, Clarke G, Mason DF, Anderson T, Richardson A, et al. Multiple sclerosis impact on employment and income in New Zealand. Acta Neurol Scand 2017 Sep;136(3):223-232. [doi: 10.1111/ane.12714] [Medline: 27891575]
18. Jayakumar P, Bozic K, Lee T. Information asymmetry: the untapped value of the patient. NEJM catalyst. 2019. URL: https://catalyst.nejm.org/doi/abs/10.1056/CAT.19.0700 [accessed 2021-08-26]

19. Reich J. Education research. Rebooting MOOC research. Science 2015 Jan 02;347(6217):34-35. [doi: 10.1126/science.1261627] [Medline: 25554779]
20. Reich J, Ruipérez-Valiente JA. The MOOC pivot. Science 2019 Jan 11;363(6423):130-131. [doi: 10.1126/science.aav7958] [Medline: 30630920]
21. Shah D. By the numbers: MOOCs in 2019. The Report. URL: https://www.classcentral.com/report/mooc-stats-2019/ [accessed 2020-02-27]
22. University of Tasmania - Menzies Institute for Medical Research. URL: https://ms.mooc.utas.edu.au/ [accessed 2021-10-20]
23. Tieman J, Miller-Lewis L, Rawlings D, Parker D, Sanderson C. The contribution of a MOOC to community discussions around death and dying. BMC Palliat Care 2018 Feb 20;17(1):31 [FREE Full text] [doi: 10.1186/s12904-018-0287-3] [Medline: 29458347]
24. Shah D. Class central's best online courses of 2019. The Report. URL: https://www.classcentral.com/report/best-free-online-courses-2019/ [accessed 2020-02-27]
25. Stiglitz JE. The contributions of the economics of information to twentieth century economics. Q J Econ 2000 Nov;115(4):1441-1478. [doi: 10.1162/003355300555015]

Abbreviations

MOOC: massive open online course
MS: multiple sclerosis
WDREC: Wicking Dementia Research and Education Centre

Edited by R Kukafka; submitted 23.06.20; peer-reviewed by R Haase, A Fakolade; comments to author 21.09.20; revised version received 07.10.20; accepted 24.05.21; published 07.12.21.

Please cite as:
Claflin SB, Campbell JA, Doherty K, Farrow M, Bessing B, Taylor BV
Evaluating Course Completion, Appropriateness, and Burden in the Understanding Multiple Sclerosis Massive Open Online Course: Cohort Study
J Med Internet Res 2021;23(12):e21681
URL: https://www.jmir.org/2021/12/e21681
doi:10.2196/21681
PMID:34878985

Original Paper

Measuring and Improving Evidence-Based Patient Care Using a Web-Based Gamified Approach in Primary Care (QualityIQ): Randomized Controlled Trial

Trever Burgon[1], PhD; Linda Casebeer[2], PhD; Holly Aasen[2], BA; Czarlota Valdenor[1], MD; Diana Tamondong-Lachica[1], MD; Enrico de Belen[1], MD; David Paculdo[1], MPH; John Peabody[1,3], MD, PhD

[1]QURE Healthcare, San Francisco, CA, United States

[2]CE Outcomes, Birmingham, AL, United States

[3]School of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
John Peabody, MD, PhD
QURE Healthcare
450 Pacific Ave
Suite 200
San Francisco, CA, 94133
United States
Phone: 1 4153213388 ext 101
Email: jpeabody@qurehealthcare.com

Abstract

Background: Unwarranted variability in clinical practice is a challenging problem in practice today, leading to poor outcomes for patients and low-value care for providers, payers, and patients.

Objective: In this study, we introduced a novel tool, QualityIQ, and determined the extent to which it helps primary care physicians to align care decisions with the latest best practices included in the Merit-Based Incentive Payment System (MIPS).

Methods: We developed the fully automated QualityIQ patient simulation platform with real-time evidence-based feedback and gamified peer benchmarking. Each case included workup, diagnosis, and management questions with explicit evidence-based scoring criteria. We recruited practicing primary care physicians across the United States into the study via the web and conducted a cross-sectional study of clinical decisions among a national sample of primary care physicians, randomized to continuing medical education (CME) and non-CME study arms. Physicians "cared" for 8 weekly cases that covered typical primary care scenarios. We measured participation rates, changes in quality scores (including MIPS scores), self-reported practice change, and physician satisfaction with the tool. The primary outcomes for this study were evidence-based care scores within each case, adherence to MIPS measures, and variation in clinical decision-making among the primary care providers caring for the same patient.

Results: We found strong, scalable engagement with the tool, with 75% of participants (61 non-CME and 59 CME) completing at least 6 of 8 total cases. We saw significant improvement in evidence-based clinical decisions across multiple conditions, such as diabetes (+8.3%, $P<.001$) and osteoarthritis (+7.6%, $P=.003$) and with MIPS-related quality measures, such as diabetes eye examinations (+22%, $P<.001$), depression screening (+11%, $P<.001$), and asthma medications (+33%, $P<.001$). Although the CME availability did not increase enrollment in the study, participants who were offered CME credits were more likely to complete at least 6 of the 8 cases.

Conclusions: Although CME availability did not prove to be important, the short, clinically detailed case simulations with real-time feedback and gamified peer benchmarking did lead to significant improvements in evidence-based care decisions among all practicing physicians.

Trial Registration: ClinicalTrials.gov NCT03800901; https://clinicaltrials.gov/ct2/show/NCT03800901

(J Med Internet Res 2021;23(12):e31042) doi:10.2196/31042

KEYWORDS

quality improvement; physician engagement; MIPS; case simulation; feedback; value-based care; care standardization; simulation; gamification; medical education; continuing education; outcome; serious game; decision-support

Introduction

Clinical practice variation is recognized as one of the most challenging problems in current practice [1,2]. Unwarranted variability in clinical practice has multiple root causes, starting with the uneven recognition and application of medical knowledge [3,4]. The sheer volume of new research, including nearly 1.4 million papers (or 1 paper every 23 seconds) posted to the National Library of Medicine's PubMed database in 2019, also makes it virtually impossible for busy practicing physicians to keep their practice up to date [5]. Our own research shows that even after case mix adjustment, practice variation within the same practice is a significant problem, characterized by standard deviations of approximately 10% [6-8]. The good news is that when knowledge and practice gaps are closed, variability declines, adoption of best practices accelerates, and patient outcomes improve [9,10]. Conversely, failing to recognize and address unwarranted variation has deleterious impacts on quality, outcomes, and value [11-16].

The challenge of reducing unwarranted clinical variation has been widely documented across care settings, clinical specialties, Merit-Based Incentive Payment System (MIPS) measures, and geographies. Easy solutions have been tried, including Continuing Medical Education (CME) and maintenance of certification (MOC), performance dashboards, and reminders; however, success has been limited [17]. For example, the most common forms of CME activities, ranging from printed information to didactic presentations and formal conferences, have shown relatively little impact on physician performance [18]. Other engagement strategies, such as multimedia approaches, multiple instructional techniques, repeated exposures, and direct feedback on care decisions, have shown better effectiveness but are difficult to scale and time-intensive for participating physicians [19].

Research using a newer approach—timely feedback on case-based decisions using validated case simulations—has been shown to lead to significant changes in actual practice in randomized controlled trials [6,20,21]. Another research stream has used the motivational aspects of gaming, real-time scoring, digital feedback, leaderboards, and serial competition, which suggests that the gaming approach provides an opportunity to enhance medical education [22-25].

The engagement tool created for this study builds on over 20 years of research using Clinical Performance and Value (CPV) patient simulations [26]. We adapted those lessons to develop a novel web-based patient-simulation platform, known as QualityIQ, which is focused on primary care providers (PCPs) and leverages the serial engagement of case-based learning in CPVs with immediate personalized evidence-based feedback and gamified peer benchmarking. QualityIQ is distinct from the standard gamification approach in that QualityIQ leverages iterative measurement, feedback, and remeasurement over multiple rounds of engagement using the CPV approach. We introduced the QualityIQ tool to PCPs to determine if serial measurement and feedback improved evidence-aligned practice decisions overall and whether it improved specific quality measures included in MIPS. After completing their cases, we determined whether receiving CME credits increased participation in this quality improvement initiative. Finally, we asked the participants directly if they expected to make changes in their actual practice setting after participating in this gamified learning approach.

Methods

Study Design

From January through March 2019, we conducted a randomized controlled study of clinical care decisions made by a national sample of PCPs managing typical primary care patients. We asked United States–based, board-certified internal medicine and family medicine physicians to care for four different types of routine primary care cases (diabetes, osteoarthritis [OA], asthma, and musculoskeletal pain). We used the novel web-based QualityIQ patient simulation tool to serially measure provider care decisions for these cases. Physicians were given real-time feedback when they completed their cases to determine the extent to which their care decisions aligned with the latest guidelines. We measured the care decisions judged to be the most critical to high-quality care, namely the workup (laboratory and imaging), diagnosis, and treatment. Gaming elements included a leaderboard for all participants and gift cards for top scores. We took advantage of the prospective design and used a coin flip methodology to randomly assign half of the participants to receive CME and the other half to not receive CME to observe whether this augmented the participation, learning, or standardization effects of serial measurement and feedback.

Physician Recruitment

From a list of over 10,000 US-based PCPs, we sent out 2000 emails to randomly selected addresses. From this group, we screened potential participants using the following enrollment criteria: (1) is board-certified in internal medicine or family medicine, (2) practices exclusively in primary care, (3) has an active panel of over 1500 patients, and (4) has 2 to 30 years of postresidency experience. In total, 202 providers were eligible, and of these, 141 agreed to participate. The 141 participants were further randomized into 1 of 2 study arms, with 68 in the non-CME control group and 73 in the intervention group that received CME with their participation. Of the 141 physicians who completed the questionnaire and enrolled in the study, 21 began the first week but did not complete their case and were subsequently dropped from the study, leaving 120 enrolled providers who completed at least one week of cases (see Table 1).

Table 1. Provider characteristics at baseline (N=120).

Characteristic	Value		*P* value
	Non-CME[a] (n=61)	CME (n=59)	
Male, n (%)	43 (70)	38 (64)	.56
Age >55 years, n (%)	29 (48)	29 (49)	.86
Region, n (%)			.05
Northeast	20 (33)	16 (27)	
Midwest	10 (16)	14 (24)	
West	8 (13)	17 (29)	
South	23 (38)	12 (20)	
Locale, n (%)			.22
Urban	27 (44)	24 (41)	
Suburban	26 (44)	32 (54)	
Rural	8 (13)	3 (5)	
Specialty, n (%)			.40
Family medicine	26 (43)	21 (36)	
Internal medicine	34 (56)	38 (64)	
Both	1 (2)	0 (0)	
Attended medical school in the United States, n (%)	49 (80)	42 (71)	.29
Practice type, n (%)			.23
Solo	15 (25)	10 (17)	
Group single-specialty	18 (30)	9 (15)	
Group multispecialty	12 (20)	21 (36)	
Hospital	5 (8)	7 (12)	
Academic	6 (10)	6 (10)	
Other	5 (8)	6 (10)	
Employed by practice, n (%)	42 (69)	51 (87)	.03
Patients seen/week, mean (SD)	101 (47)	87 (33)	.07
Receive quality bonus, n (%)	35 (57)	30 (51)	.58
Participation in CMS[b] quality payment programs, n (%)			
MIPS[c]	27 (44)	20 (33.9)	.27
APM[d]	9 (15)	7 (12)	.79
Other	3 (5)	3 (5)	.97
None	18 (30)	12 (20)	.29
Number of rounds of participation, mean (SD)	4.5 (3.2)	6.1 (2.7)	.003
Participated in ≥6 rounds, n (%)	29 (48)	40 (66)	.045

[a]CME: Continuing Medical Education.

[b]CMS: Centers for Medicare & Medicaid Services.

[c]MIPS: Merit-Based Incentive Payment System.

[d]APM: Advanced Payment Model.

QualityIQ Patient Simulation Cases

We created 8 fully automated QualityIQ case simulations and uploaded these cases onto the Qualtrics platform [27]. Each case included evidence-based feedback delivered in real time as physicians progressed through various workup, diagnosis, and treatment decisions. Each case was designed to be completed on a smartphone, tablet, or computer in less than 10 minutes. Each week, all participants cared for the same case.

The 8 cases were developed as pairs of typical cases seen by PCPs in four areas: diabetes, OA, asthma, and pain management (see Table S1 in Multimedia Appendix 1). While each case was unique and required different treatment decisions based on each patient's presenting symptoms and risk factors, many care decisions were featured in multiple cases (see Table S2 in Multimedia Appendix 2). For example, we included decisions directly related to Medicare 2019 MIPS measures, such as addressing poor hemoglobin A_{1c} control (>9%). We also included general measures that cut across multiple conditions, such as zoster vaccination. By having multiple related scoring items in multiple cases, we were able to track changes over time.

QualityIQ Scoring and Gamification

The PCPs completed 1 case per week, with weekly email reminders to notify them when the next case opened. Each weekly case consisted of 8 to 10 multiple choice questions covering workup, diagnosis, management, and follow-up decisions, and each question had explicit evidence-based scoring criteria. After each question, physicians received real-time feedback on their care decisions, including the appropriateness of their decision, recommended alternative decisions, and supporting evidence-based references for the preferred care path.

At the end of each week, participants received a detailed score report that included a summary of key evidence-based recommendations for their case, their personal score in the case, and how their care compared to that of their peers. At the start of the study, all participants chose a pseudonym so they could track their scores relative to their peers on a leaderboard that was updated weekly. The top scores in each weekly case were awarded a US $20 electronic gift card from Amazon. The study was completed after the close of the final case.

Statistical Analysis

The primary outcomes were to measure evidence-based care scores within each case, adherence to MIPS measures, and practice variability among the PCPs caring for the same patient. We were especially keen to determine if the physicians improved their scores on these measures after serial measurements. We also investigated if the availability of CME credit had any effect on participation or performance. Lastly, we asked the participants for their appraisal of the usefulness of the tool in their practice.

For descriptive comparisons between the 2 study arms, we used the chi-square test for significance. To determine significance across cases, we normalized the scores to percentages; a score of 100% indicated that the PCP made all the correct evidence-based decisions without any incorrect decisions, with a possible score of less than 0% if the PCP made more incorrect than correct decisions. We compared these normalized quality-of-care scores across the cases using either multivariate linear regression or the Student *t* test to measure improvements in overall and domain quality of care scores. We also performed an equality of variances test to test for homogeneity of the overall scores. All analyses were conducted in Stata 14.2 (StataCorp LLC).

Ethics Approval and Consent to Participate

This study was conducted in accordance with ethical standards, approved by the Advarra Institutional Review Board, Columbia, Maryland, and listed on ClinicalTrials.gov (NCT03800901). Informed consent was obtained through electronic signatures from all participants.

Results

Physician Demographics

Of the 120 participants in the study, more than two-thirds were male, and 72 (60%) were board certified in internal medicine. Among the demographics and practice characteristics listed in Table 1, we found no significant differences between the two groups except that the CME group had a higher percentage of providers who were employed by their practice (86.4% vs 68.9%; *P*=.03).

All 120 participants cared for one QualityIQ patient in the first week of the project. In the second week, 91 (76%) of the 120 participants completed their second case. After week 2, participation stabilized, with only modest decreases from weeks 3 to 8. Of the 91 participants who completed at least 2 cases, 68 (75%) went on to complete at least 6 of the 8 weekly cases. 58 (48%) physicians completed all 8 cases, and 79 (66%) participated in at least half (n=4) of the cases. When we compared the first week scores between providers who completed 8 weeks of the study to those who only completed the first week of the study, we found no significant difference in their scores (*P*=.37).

The ability to earn CME did not affect recruitment rates. However, once enrolled, those eligible for CME credits completed an average of 1.6 more cases in the project than their non-CME peers (*P*=.003) and were more likely to participate in at least 6 of the weekly rounds (40 of 61, 66.1%, vs 28 of 59, 47.5%; *P*=.045).

In aggregate, female physicians performed significantly better than their male counterparts (+3.1%, *P*=.02), and family medicine diplomates performed better than internal medicine providers (+3.2%, *P*=.008) (see Table 2). We saw no significant difference in overall scores by age, with providers aged over 55 years scoring a nonsignificant 0.7% lower than their younger counterparts (*P*=.56). In our study, those practicing in multispecialty group practices (+6.5%) and those practicing in the Midwest region (+8.1%) scored significantly higher (*P*<.001 for both). PCPs who participated in 6 or more weeks of QualityIQ cases had higher average quality scores than those who participated in 5 or fewer weeks (+5.2%, *P*=.04). However, providers who were randomized into the CME arm of the study did not perform better than those in the non-CME arm (+0.5%, *P*=.84).

Table 2. Multivariate linear regression analysis of total QualityIQ scores (as percentages of the maximum score).

Characteristic	Coefficient	*P* value
Male sex	–3.1	.02
Internal medicine physician	–3.2	.008
Age >55 years	–0.7	.56
US-trained physician	–0.1	.97
Midwest region	8.1	<.001
Suburban locale	2.0	.12
Multispecialty group practice	6.5	<.001
Academic practice	4.9	.01
Received quality bonus	0.8	.50
Case type[a]		
Osteoarthritis	–10.1	<.001
Asthma	–6.9	<.001
Pain	–8.4	<.001
Second case of type	6.4	<.001
Participation ≥6 rounds	5.2	.03
CME[b]	0.5	.84
Participation ≥6 rounds * CME	–0.6	.84
Constant	74.7	<.001

[a]Reference case type: diabetes.

[b]CME: Continuing Medical Education.

Reduction in Variability of Care

Overall, we found a 9.2% reduction in variation between the first and second cases for each case type (*P*=.07). There were different levels of reduction by case type. For example, the relative standard deviation decreased by 37.0% (*P*<.001) in the diabetes cases. When we disaggregated this further, the decreased variation was split fairly evenly between the treatment domain, where the standard deviation decreased by 34.1% (*P*<.001), and a 33.1% reduction was observed in the diagnostic domain (*P*<.001). Variation decreased between the OA and asthma case pairs, but not between the pain cases. In the OA cases, we found a 12.5% relative decrease in variation (*P*=.14), and in the asthma cases, we saw a 15.9% decrease (*P*=.08).

Quality of Care Improvement Overall and by Case

In the first week of the project, the average score was 77%. When we compared changes in scores among the different case pairs over time (ie, diabetes, OA, asthma, and pain; see Table 3), we found that providers performed 1 to 10 percentage points better in the second case compared to the first. These improvements were statistically significant for patients with diabetes, OA, and asthma but not for the pain case pair. When we looked at the mean scores for the OA and asthma case pairs, we saw a significant increase in the mean scores (Table 3), with the OA case scores improving by 7.6% (*P*=.003) and the asthma scores improving by 10.7% (*P*<.001).

Table 3. Summary of QualityIQ results.

Case type and week	Maximum total score	All providers			*P* value
		n	Mean total score	Percentage of maximum score, mean (SD)	
Diabetes mellitus					<.001
1	350	120	272	77.6 (14.6)	
7	350	74	301	85.9 (9.2)	
Osteoarthritis					.003
2	270	85	185	68.5 (16.5)	
4	330	76	251	76.1 (14.6)	
Asthma					<.001
3	260	81	184	70.8 (15.1)	
6	350	65	285	81.5 (12.7)	
Pain					.73
5	320	72	236	73.7 (12.8)	
8	310	65	231	74.5 (15.9)	

Improvement in MIPS-Related Measures

We found that baseline performance on the specific MIPS-related scoring items ranged from 21% for screening and brief counseling for unhealthy alcohol use to 100% for prescribing high blood pressure medication (Table 4). In comparing the two study arms, as well as family medicine versus internal medicine providers, we found no overall differences between the two groups. There were instances of significance, which might be expected with a subanalysis; for example, the CME arm was more than twice as likely (odds ratio [OR] 2.2, 95% CI 1.2-3.8) to order pneumococcal immunization than the non-CME arm, and internal medicine providers were half as likely (OR 0.5, 95% CI 0.3-0.8) to screen for depression.

Table 4. Change in Merit-Based Incentive Payment System (MIPS) measures over time.

MIPS measure, category, and name	Ordering, normalized percentage	*P* value
1. Treatment Diabetes: Hemoglobin A_{1c} (HbA_{1c}) Poor Control (>9%)		<.001
Week 1	63	
Week 7	96	
110. Preventive Care and Screening: Influenza Immunization		.58
Week 1	96	
Week 2	95	
Week 3	96	
Week 4	96	
Week 6	100	
111. Preventive Care and Screening: Pneumococcal Vaccination Status for Older Adults		.34
Week 1	72	
Week 4	80	
Week 7	71	
113. Preventive Care and Screening: Colorectal Cancer Screening		.72
Week 1	92	
Week 4	88	
Week 7	90	
117. Treatment: Diabetes: Eye Exam		<.001
Week 1	74	
Week 7	96	
126. Treatment: Diabetes Mellitus: Diabetic Foot and Ankle Care, Peripheral Neuropathy – Neurological Evaluation		.07
Week 1	83	
Week 7	92	
134. Preventive Care and Screening: Screening for Depression and Follow-Up Plan		<.001
Week 1	84	
Week 2	71	
Week 3	70	
Week 4	96	
Week 6	83	
Week 7	96	
Week 8	95	
226. Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention		.31
Week 5	93	
Week 6	97	
236. Treatment: Controlling High Blood Pressure		.04
Week 1	58	
Week 2	58	
Week 4	67	
Week 7	77	
309. Preventive Care and Screening: Cervical Cancer Screening		.42

MIPS measure, category, and name	Ordering, normalized percentage	*P* value
Week 3	84	
Week 5	79	
398. Treatment: Optimal Asthma Control		.048
Week 3	99	
Week 6	98	
431. Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling		<.001
Week 5	21	
Week 6	50	
Week 7	53	
Week 8	55	
438. Treatment: Statin Therapy for the Prevention and Treatment of Cardiovascular Disease		.46
Week 1	92	
Week 2	78	
Week 7	95	
444. Treatment: Medication Management for People with Asthma		<.001
Week 3	62	
Week 6	95	
474. Preventive Care and Screening: Zoster (Shingles) Vaccination		<.001
Week 1	78	
Week 2	58	
Week 4	95	
Week 5	80	
Week 7	77	

Not surprisingly, measures with baseline performance above 80% showed minimal improvement in subsequent cases. These high-performing measures appeared to be well-established items in primary care practice, such as influenza immunization, colorectal cancer screening, and statin therapy. By contrast, MIPS-related scoring items with baseline performance <80% demonstrated strong and statistically significant improvements through serial measurement and feedback. Notable examples include a 22% increase in diabetic eye examination referrals (P<.001), an 11% increase in depression screening (P<.001), a 19% increase in appropriate identification of blood pressure goals (P=.04), and a 33% increase in evidence-based asthma medication recommendations (P<.001). Pneumococcal vaccination for older adults was the lone exception; it started at 72% in the baseline case but did not demonstrate a significant improvement (P=.34) in the 3 subsequent cases that included this care decision.

Physician Survey Results

After the 8 weeks of the project were complete, we asked the physicians about the usefulness of this approach. Of the 120 participants, 62 responded (a 52% response rate). 89% rated the overall quality of the material as good or excellent; 76% reported that they plan to do something in differently in their practice based on what they learned in the cases and the feedback. In addition, participants rated their satisfaction with the gamified weekly leaderboard at 4.1 out of 5.0 on a Likert scale. Importantly, the participants gave the project a net promoter score (NPS) of 59, indicating a strong preference that they would recommend the program to their primary care colleagues.

Discussion

Principal Results

Finding effective tools that reduce the variation in clinical practice has been challenging. Traditional CME tools have not shown knowledge retention, and scalable engagement has proven difficult to implement [17,18]. Recent studies have shown that active case-based learning and more interactive techniques, gamification, and deliberate practice show promise in boosting physician engagement, enhancing mastery learning, and improving clinical care quality [23,28-30]. Reducing practice variation and increasing the quality of care patients receive may be most urgently needed in primary care, where the high volume of patients and large breadth of conditions managed are particularly manifest.

This study, which introduced the QualityIQ tool to reduce practice variation, had a few notable findings. Participation rates were high over multiple exposures, with 66% of participants

completing at least half of the weekly cases. This is significant because participation was voluntary and offered without any emoluments beyond gamification and recognition on an anonymous leaderboard. The findings also suggest that physicians are interested in efficient and engaging tools that help providers stay abreast of the latest guidelines. Interestingly, the availability of CME and MOC credits had no impact on recruitment into the activity or on performance in the cases, although once a participant joined, they were more likely to complete more cases if they were randomized to CME. We believe that the proliferation of web-based CME opportunities means that fewer physicians need to seek out CME opportunities.

The most significant finding from our study is that iterative measurement, feedback, and remeasurement over multiple rounds of engagement led to significant reductions in care variation (variation reduction by case type: asthma: –15.9%, P=.08; osteoarthritis: –12.5%, P=.14; diabetes: –37.0%, P<.001). There were also broad-based improvements in care decisions from one case to the other (by case type: asthma: +15.1%, P<.001; osteoarthritis: +11.1%, P=.003; diabetes: +10.7%, P<.001). There was no decreased variation or improvement in the pain management cases, which we attribute to two factors: (1) the pain case pairs were too clinically dissimilar (headache and low back pain), and (2) the established clinical guidelines for pain management are less robust that for the other case types. This lack of findings in the pain case type is a strong indicator that the improvements seen in the other case pairs was not simply a "learning effect" bias, wherein participants simply became accustomed to the format.

The specific MIPS-measured care decisions were assessed across multiple cases and also showed improvements with multiple exposures. These improvements extended across preventive and treatment clinical areas, and the measures with the lowest baseline performance showed the strongest improvements. MIPS measures that were adhered to less than 80% of the time at baseline specifically improved between 11% and 33% (P<.05). These may be especially important for commonly overlooked items (eg, depression screening) and new items where the guidelines have changed recently (eg, zoster vaccination). Pneumococcal immunization was the outlier, not improving over time from its baseline performance of approximately 70%. This may reflect disagreements with the guideline-based recommendations, which were subsequently updated by the US Centers for Disease Control and Prevention Advisory Committee on Immunization Practices in June 2019, after completion of our data collection [31].

Practice improvement tools only have impact if they are welcome and adopted. Accompanying these improvements, we found corroborating self-reports of practice change among the physician participants and enthusiastic reception of the tool, with an NPS of 59. The NPS is indicative of a user or client's experience. Users are first asked to rate how likely they are to recommend a service to others. The NPS is then determined by determining the difference between the percentage of promoters (satisfied clients who give a score of 9-10) and the percentage of detractors (dissatisfied clients who gave a score of 0-6). A score above zero can be considered a good score, meaning there are more promoters than detractors, and a score above 50 is considered excellent [32]. In addition, the gamified leaderboard allowing peer-to-peer comparisons using pseudonyms was well received by participants. Another noteworthy finding, given concerns that web-based or digital tools may not reach older physicians, is that physicians over the age of 55 years performed as well as other providers, suggesting that the approach may be broadly applicable to practicing PCPs at various stages of their career.

Limitations

There are limitations to this validation study that are worth noting. Although an impressive 76% of participants reported making changes to their practice based on their participation in the QualityIQ cases and feedback, the study was not designed to interrogate practice or patient-level records to validate these improvements. This important work will be left to future studies. In addition, this 8-week curriculum covered a number of cases typically seen in primary care, but it did not include an exhaustive range of high-priority topics. This could be addressed through longer-term studies, potentially in partnership with health systems or physician groups. The project was designed to simulate actual practice decisions through simulations rather than create a fully validated examination. As such, questions formulated around areas of clinical relevance were tied to typical practice patterns. Psychometric validation of the questions was not performed but could be a priority for future academic research applications of the tool. In addition, although PCPs play a critically important role in quality improvement, there are significant opportunities to improve care quality among specialist physicians, medical trainees, nurses, and other health care professionals. Future work will elucidate the impact of this engagement model in these other settings.

Conclusions

In recognition of the vital role of primary care, multiple programs from government and nongovernment agencies have prioritized primary care practice improvement as essential to care transformation efforts to improve care quality and value. In this study, we have shown that short case simulations delivering real-time personalized feedback and gamified peer benchmarking are very well received by practicing primary care physicians and lead to significant improvements in evidence-based care decisions. Importantly, as the QualityIQ scores increased, the unwarranted variation between providers decreased, which is a "holy grail" in efforts to build high-quality, high-reliability primary care networks. As a web-based, scalable engagement tool, this model may be of interest to health systems, payers, policy makers, patient advocacy groups, and life science companies looking to collaborate with providers in practice change efforts to improve the quality, value, and consistency of care.

Data Availability

The data sets used to support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

No funding was received for this work.

Authors' Contributions

JP, LC, and TB designed the framework and planned the study. CV, DTL, and ED planned and developed the cases used in the study. TB, HA, and DTL supervised the study. JP, DP, and ED analyzed the data. TB, DP, and JP drafted the manuscript. All authors contributed to finalizing and reviewing the manuscript.

Conflicts of Interest

QURE, LLC owns the intellectual property used to prepare the cases and collect the data. JP is the owner of QURE, LLC. TB, CV, DTL, EDB, and DP are employees of QURE Healthcare.

Multimedia Appendix 1
QualityIQ case summaries.
[DOCX File , 16 KB - jmir_v23i12e31042_app1.docx]

Multimedia Appendix 2
Merit-Based Incentive Payment System measures by case.
[DOCX File , 19 KB - jmir_v23i12e31042_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 716 KB - jmir_v23i12e31042_app3.pdf]

References

1. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ 2011 Mar 17;342:d1513-d1513. [doi: 10.1136/bmj.d1513] [Medline: 21415111]
2. Atsma F, Elwyn G, Westert G. Understanding unwarranted variation in clinical practice: a focus on network effects, reflective medicine and learning health systems. Int J Qual Health Care 2020 Jun 04;32(4):271-274 [FREE Full text] [doi: 10.1093/intqhc/mzaa023] [Medline: 32319525]
3. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec 16;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
4. Sutherland K, Levesque J. Unwarranted clinical variation in health care: definitions and proposal of an analytic framework. J Eval Clin Pract 2020 Jun 28;26(3):687-696 [FREE Full text] [doi: 10.1111/jep.13181] [Medline: 31136047]
5. PubMed. National Library of Medicine. URL: https://tinyurl.com/3m7ey7jp [accessed 2021-01-26]
6. Bergmann S, Tran M, Robison K, Fanning C, Sedani S, Ready J, et al. Standardising hospitalist practice in sepsis and COPD care. BMJ Qual Saf 2019 Oct 20;28(10):800-808. [doi: 10.1136/bmjqs-2018-008829] [Medline: 30894422]
7. Weems L, Strong J, Plummer D, Martin J, Zweng TN, Lindsay J, et al. A quality collaboration in heart failure and pneumonia inpatient care at Novant Health: standardizing hospitalist practices to improve patient care and system performance. Jt Comm J Qual Patient Saf 2019 Mar;45(3):199-206. [doi: 10.1016/j.jcjq.2018.09.005] [Medline: 30391372]
8. Oravetz P, White CJ, Carmouche D, Swan N, Donaldson J, Ruhl R, et al. Standardising practice in cardiology: reducing clinical variation and cost at Ochsner Health System. Open Heart 2019 Mar 22;6(1):e000994 [FREE Full text] [doi: 10.1136/openhrt-2018-000994] [Medline: 30997137]
9. Yurso M, Box B, Burgon T, Hauck L, Tagg K, Clem K, et al. Reducing unneeded clinical variation in sepsis and heart failure care to improve outcomes and reduce cost: a collaborative engagement with hospitalists in a multistate system. J Hosp Med 2019 Jun 19;14(9):E1-E6 [FREE Full text] [doi: 10.12788/jhm.3220] [Medline: 31251162]
10. Peabody JW, Quimbo S, Florentino J, Shimkhada R, Javier X, Paculdo D, et al. Comparative effectiveness of two disparate policies on child health: experimental evidence from the Philippines. Health Policy Plan 2017 May 01;32(4):563-571 [FREE Full text] [doi: 10.1093/heapol/czw179] [Medline: 28110265]
11. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. N Engl J Med 2003 Jun 26;348(26):2635-2645. [doi: 10.1056/NEJMsa022615] [Medline: 12826639]
12. McCulloch P, Nagendran M, Campbell WB, Price A, Jani A, Birkmeyer JD, et al. Strategies to reduce variation in the use of surgery. Lancet 2013 Sep;382(9898):1130-1139. [doi: 10.1016/s0140-6736(13)61216-7]

13. Davis P, Gribben B, Lay-Yee R, Scott A. How much variation in clinical activity is there between general practitioners? A multi-level analysis of decision-making in primary care. J Health Serv Res Policy 2002 Oct 23;7(4):202-208. [doi: 10.1258/135581902320432723] [Medline: 12425779]
14. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. BMJ 2002 Oct 26;325(7370):961-964 [FREE Full text] [doi: 10.1136/bmj.325.7370.961] [Medline: 12399352]
15. Krumholz HM, Normand ST, Spertus JA, Shahian DM, Bradley EH. Measuring performance for treating heart attacks and heart failure: the case for outcomes measurement. Health Aff (Millwood) 2007 Jan;26(1):75-85. [doi: 10.1377/hlthaff.26.1.75] [Medline: 17211016]
16. Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena AB. Variation in physician spending and association with patient outcomes. JAMA Intern Med 2017 May 01;177(5):675-682 [FREE Full text] [doi: 10.1001/jamainternmed.2017.0059] [Medline: 28288254]
17. Davis D. Does CME work? An analysis of the effect of educational activities on physician performance or health care outcomes. Int J Psychiatry Med 1998 Apr 01;28(1):21-39. [doi: 10.2190/ua3r-jx9w-mhr5-rc81]
18. Bloom BS. Effects of continuing medical education on improving physician clinical care and patient health: a review of systematic reviews. Int J Technol Assess Health Care 2005 Aug 04;21(3):380-385. [doi: 10.1017/s026646230505049x] [Medline: 16110718]
19. O'Neil KM, Addrizzo-Harris DJ, American College of Chest Physicians Health and Science Policy Committee. Continuing medical education effect on physician knowledge application and psychomotor skills: effectiveness of continuing medical education: American College of Chest Physicians evidence-based educational guidelines. Chest 2009 Mar;135(3 Suppl):37S-41S. [doi: 10.1378/chest.08-2516] [Medline: 19265074]
20. Burgon TB, Cox-Chapman J, Czarnecki C, Kropp R, Guerriere R, Paculdo D, et al. Engaging primary care providers to reduce unwanted clinical variation and support ACO cost and quality goals: a unique provider-payer collaboration. Popul Health Manag 2019 Aug;22(4):321-329. [doi: 10.1089/pop.2018.0111] [Medline: 30328782]
21. Richards JM, Burgon TB, Tamondong-Lachica D, Bitran JD, Liangco WL, Paculdo DR, et al. Reducing unwarranted oncology care variation across a clinically integrated network: a collaborative physician engagement strategy. JOP 2019 Dec;15(12):e1076-e1084. [doi: 10.1200/jop.18.00754]
22. Harrison R, Hinchcliff RA, Manias E, Mears S, Heslop D, Walton V, et al. Can feedback approaches reduce unwarranted clinical variation? A systematic rapid evidence synthesis. BMC Health Serv Res 2020 Jan 16;20(1):40 [FREE Full text] [doi: 10.1186/s12913-019-4860-0] [Medline: 31948447]
23. Peabody J, Oskombaeva K, Shimarova M, Adylbaeva V, Dzhorupbekova K, Sverdlova I, et al. A nationwide program to improve clinical care quality in the Kyrgyz Republic. J Glob Health 2020 Dec 15;10(2):020418 [FREE Full text] [doi: 10.7189/jogh.10.020418] [Medline: 33110578]
24. Lai AKH, Noor Azhar AMB, Bustam AB, Tiong XT, Chan HC, Ahmad RB, et al. A comparison between the effectiveness of a gamified approach with the conventional approach in point-of-care ultrasonographic training. BMC Med Educ 2020 Aug 12;20(1):263 [FREE Full text] [doi: 10.1186/s12909-020-02173-7] [Medline: 32787921]
25. Nevin CR, Westfall AO, Rodriguez JM, Dempsey DM, Cherrington A, Roy B, et al. Gamification as a tool for enhancing graduate medical education. Postgrad Med J 2014 Dec 28;90(1070):685-693 [FREE Full text] [doi: 10.1136/postgradmedj-2013-132486] [Medline: 25352673]
26. Peabody JW, Luck J, Glassman P, Jain S, Hansen J, Spell M, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. Ann Intern Med 2004 Nov 16;141(10):771-780 [FREE Full text] [doi: 10.7326/0003-4819-141-10-200411160-00008] [Medline: 15545677]
27. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2021-11-25]
28. Mehta N, Geissel K, Rhodes E, Salinas G. Comparative effectiveness in CME: evaluation of personalized and self-directed learning models. J Contin Educ Health Prof 2015;35 Suppl 1:S24-S26. [doi: 10.1002/chp.21284] [Medline: 26115238]
29. Cervero RM, Gaines JK. The impact of CME on physician performance and patient health outcomes: an updated synthesis of systematic reviews. J Contin Educ Health Prof 2015;35(2):131-138. [doi: 10.1002/chp.21290] [Medline: 26115113]
30. Kerfoot BP, Turchin A, Breydo E, Gagnon D, Conlin PR. An online spaced-education game among clinicians improves their patients' time to blood pressure control: a randomized controlled trial. Circ Cardiovasc Qual Outcomes 2014 May;7(3):468-474. [doi: 10.1161/circoutcomes.113.000814]
31. Matanock A, Lee G, Gierke R, Kobayashi M, Leidner A, Pilishvili T. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged ≥65 years: updated recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 2019 Nov 22;68(46):1069-1075 [FREE Full text] [doi: 10.15585/mmwr.mm6846a5] [Medline: 31751323]
32. Amaresan S. What is a good net promoter score? HubSpot. 2019 Oct 02. URL: https://blog.hubspot.com/service/what-is-a-good-net-promoter-score [accessed 2020-12-06]

Abbreviations

CME: continuing medical education

CPV: Clinical Performance and Value
MIPS: Merit-Based Incentive Payment System
MOC: maintenance of certification
NPS: net promoter score
OA: osteoarthritis
OR: odds ratio
PCP: primary care provider

Edited by R Kukafka; submitted 07.06.21; peer-reviewed by R Shimkhada, J Willig, K Glover; comments to author 04.08.21; revised version received 21.09.21; accepted 29.10.21; published 23.12.21.

Please cite as:
Burgon T, Casebeer L, Aasen H, Valdenor C, Tamondong-Lachica D, de Belen E, Paculdo D, Peabody J
Measuring and Improving Evidence-Based Patient Care Using a Web-Based Gamified Approach in Primary Care (QualityIQ): Randomized Controlled Trial
J Med Internet Res 2021;23(12):e31042
URL: https://www.jmir.org/2021/12/e31042
doi:10.2196/31042
PMID:34941547

Original Paper

Drinking and Social Media Use Among Workers During COVID-19 Pandemic Restrictions: Five-Wave Longitudinal Study

Atte Oksanen[1*], DrSocSci; Reetta Oksa[1*], MSocSci; Nina Savela[1*], MSocSci; Magdalena Celuch[1*], MA; Iina Savolainen[1*], DrSocSci

Tampere University, Faculty of Social Sciences, Tampere, Finland

[*]all authors contributed equally

Corresponding Author:
Atte Oksanen, DrSocSci
Tampere University
Faculty of Social Sciences
Kalevantie 5
Tampere, 33014
Finland
Phone: 358 503187279
Email: atte.oksanen@tuni.fi

Abstract

Background: The COVID-19 pandemic restricted everyday life during 2020-2021 for many people worldwide. It also affected alcohol consumption patterns and leisure activities, including the use of social media.

Objective: The aim of this study was to analyze whether social media use predicts increased risky drinking over time and during the COVID-19 pandemic restrictions in particular.

Methods: This 5-wave longitudinal survey study, based on a nationwide sample of workers, was conducted in Finland in 2019-2021. A total of 840 respondents (male: 473/840, 56.31%; age range 18-64 years; mean age 43.90, SD 11.14 years) participated in all 5 waves of the study. The outcome variable was risky drinking, measured using the 3-item Alcohol Use Disorders Identification Test (AUDIT-C). Multilevel linear hybrid modeling enabled the investigation of both within-person and between-person effects. Predictors included social media use and communication, involvement in social media identity bubbles, psychological distress, and remote working. Controls included sociodemographic factors and the Big Five personality traits.

Results: Increased involvement in social media identity bubbles was associated with an increase in risky drinking behavior. Of all social media platforms examined, online dating app use was associated with riskier use of alcohol over time during the COVID-19 crisis. Daily social media communication with colleagues about nonwork topics was associated with risky drinking. Female gender, younger age, university education, nonindustrial occupational field, conscientiousness, agreeableness, and neuroticism were associated with lower levels of risky drinking.

Conclusions: Social media use during a pandemic carries some risks for alcohol consumption. Involvement in social media identity bubbles and online dating are risk factors for excessive drinking during the COVID-19 pandemic.

(J Med Internet Res 2021;23(12):e33125) doi:10.2196/33125

KEYWORDS

excessive drinking; alcohol; COVID-19; social media; remote work; psychological distress; distress; pattern; trend; prediction; survey; app; risk

Introduction

The COVID-19 pandemic has remarkably changed everyday life for people worldwide. Restrictions placed on social gatherings, public events, bars, and restaurants are likely to affect peoples' drinking habits [1,2]. Researchers have argued that the COVID-19 crisis has put home drinking in focus more than ever before [3,4]. Recent studies have reported increased drinking during the COVID-19 pandemic [5-8]; however, few cross-sectional studies have reported both a decrease as well as no change in drinking [9-11]. Thus far, only a few studies have been based on longitudinal samples [1,6,7,12,13]. Additionally, it seems that individuals react to the COVID-19 crisis differently, and those facing emotional worry and distress are

especially at risk for increased drinking [1,9,14-18]. There might be specific risk factors associated with increased drinking during the COVID-19 pandemic, and the role of social media in drinking, in particular, calls for attention.

Different forms of social media have become a central part of communication and interaction among people across the world [19-21]. Social media has evolved through a transformation from Web 1.0 society to Web 2.0 society, with web-based global services enabling individuals, communities, and organizations to engage, produce, co-create, modify, and share individualized user-generated content, as well as interact, collaborate, and manage their social networks on the internet [21,22]. Social media has numerous advantages, such as quick transmission of information and fostering social connectedness [23,24]. However, social media use also has downsides, and previous research has established a relationship between social media use and higher alcohol consumption [25,26], which has been investigated particularly among young users and college students [27-33]. Furthermore, it seems that some social media apps are more popular among risk drinkers. For example, studies on Tinder users have shown that they drink more than nonusers do [34-36]. Dating app users are also more likely to engage in hookups involving alcohol use [37,38].

The underlying mechanisms between social media use and related alcohol use are not entirely known, although previous studies suggest that perceived drinking norms and exposure to alcohol content and marketing might be possible factors [25,39-41]. A meta-analysis showed that exposure to and engagement with alcohol-related content on social media were associated with higher self-reported drinking and alcohol problems among adolescents and young adults [42]. Individuals' attitudes and positive alcohol-related cognitions are motivational determinants of alcohol use [39]. Moreover, social media has been argued to be a source for social influence altering peer-drinking norms [32]. Drinking-related content on social media has become common for some users during the pandemic, and it has potential to lead to increased drinking [43].

Research on social media identity bubbles (ie, social cliques) shows that people are drawn toward similar-minded individuals with whom they identify strongly and share information [21,44,45]. Thus far, identity bubbles based on drinking during the COVID-19 pandemic have not been investigated. It is likely that in the absence of regular social activities and places to drink, people engage in drinking while interacting with their friends and colleagues on the internet. Social media is known to induce the development and duration of web-based social drinking events for normal and risky drinkers, enabling large gatherings of groups regardless of their location and prior familiarity [25]. Social media apps may promote users to find other similar-minded individuals to drink with at private parties, although this practice is not encouraged by health officials during the pandemic [46].

There are likely to be major differences in drinking during the COVID-19 pandemic at the individual level. Although the ongoing pandemic has been reported to have a negative impact on the mental well-being of people in Europe [47,48], it seems that mental health of the Finnish working population did not decline dramatically during the COVID-19 crisis, probably due to the relatively mild course of the pandemic in Finland [49]. Nevertheless, previous studies have shown that, for many people, lockdowns and changes in work conditions, with the expansion of remote work, have been associated with increased psychological distress [1,24,50,51]. In turn, psychological distress has been linked to an upsurge in alcohol use [1,52,53].

According to pre-pandemic studies, younger people and men are more likely to be risky drinkers [54-56]. Some studies have found similar connections during the COVID-19 pandemic for men [8,12,17,57] and people under 30 years of age [1]. However, female gender and younger age have been associated with both an increase and decrease in drinking [5,8,9,17,58]. The rise in drinking also varies across industries; for example, business, communication, and technology sector employees, as well as public administration employees, have reported increased drinking [1]. Higher education has been linked with increased drinking [58]. Poor financial situation has been associated with increased drinking [16,59], but also with reduced drinking [9]. A European study found that changes in drinking patterns were associated with financial stress, particularly among high-income groups—a decrease in drinking was observed among those reporting no financial stress and an increase in drinking was noted among those reporting financial distress [10]. People with conscientious personality characteristics were less likely to increase drinking during the COVID-19 crisis [1].

This longitudinal study focused on analyzing risky drinking behaviors before and during the COVID-19 pandemic restrictions in Finland. The first COVID-19 case in Finland was reported in January 2020, and the country reacted to the rising worldwide crisis in March 2020 by placing several restrictions: large public events were cancelled and recommendations for remote work were implemented [60]. On March 16, 2020, a national state of emergency was declared. Following the new emergency legislation, bars and nightclubs were forced to close, and restaurants were only allowed to sell food and low-alcohol beverages (eg, beer) to go. Finland does not allow alcohol to be delivered to homes, but monopoly stores and supermarkets were open as usual. The state of emergency was lifted in mid-June 2020; however, new restrictions were implemented again in October 2020. These restrictions concerned large public events, limiting the amount of people allowed to gather and the hours of operation of bars, nightclubs, and restaurants. The emergence of the new Delta variant of SARS-CoV-2 presented challenges to the existing guidelines. Subsequently, a state of emergency was declared again in the beginning of March 2021, leading to the closure of bars and nightclubs, and restaurants were only allowed to sell food and low-alcohol beverages to go. The second state of emergency lasted until April 27, 2021. The restrictions for bars, nightclubs, restaurants, and cultural events were gradually eased up during summer and autumn 2021. The recommendation for remote work was in place until October 15, 2021.

This longitudinal study aimed to analyze risky drinking behaviors before and during the COVID-19 pandemic restrictions in Finland. The research questions were (1) How did alcohol use change during the COVID-19 pandemic restrictions? (2) Do social media–related factors predict alcohol

use over time? (3) Are psychological and social factors, such as psychological distress, remote working, and personality, associated with drinking?

Methods

Participants

The longitudinal *Social Media at Work in Finland Survey* study started in Finland in mid-March 2019, and data were collected at 6-month intervals. A total of 5 waves were collected from Finnish employees. The baseline survey data collection commenced in mid-March 2019 (time point 1 [T1]; N=1817), with a response rate of 28.3%. The same participants were contacted for the follow-up surveys in mid-September 2019 (T2; n=1318), mid-March 2020 (T3; n=1081), mid-September 2020 (T4; n=1152), and mid-March 2021 (n=1018). The third survey was sent only to those who had responded to the second survey, whereas the fourth and fifth surveys were sent to all original respondents. Of the original 1817 survey respondents, 840 (46.23%) participated in all 5 surveys.

The survey was designed to represent the Finnish working population. The survey was conducted in Finnish, and responses were collected from all areas of mainland Finland. All participants were working at T1, and they were from various major occupational fields (see Multimedia Appendix 1 for details). Comparison of T1 participants to official statistics on Finnish workers provided by Statistics Finland showed that data collection was successful. We found no major deviations in sociodemographic factors, including age, gender, and geographical area [61-63]. The sample was similar to the official census data in other ways as well. Although official census figures are not available for the working population only, the proportion of excessive drinking in our sample was close to national statistics data for the 20-64 age group (ie, 30.35% male and 19.59% female participants in our sample vs 39% men and 24% women in the national population) [64]. Among the Finnish population of the 16-89 years age group, 58% used Facebook (vs 79.31% in our T1 sample), 39% used Instagram (vs 52.01% in our T1 sample), and 13% used Twitter (24.49% in our T1 sample) [65]. These differences are explained by the fact that our T1 sample included only workers aged 18 to 65 years.

Nonresponse analysis between those who responded to all 5 study waves (n=840) and those who responded only to some waves (n=977) showed that the final sample included slightly older respondents and more male participants. However, we found no differences in other variables included in this study, such as risk drinking, between those who had participated in all 5 surveys and others. In other words, dropout from the survey is not associated with our main points of interest. Our final sample is very close to the official census figures provided by Statistics Finland in terms of mean age (43.90 years in our sample vs 41.81 years in the Finnish population), gender (56.31% male participants in our sample vs 51.85% in the Finnish population), and university-level education (47.74% in our sample vs 42.46% in the Finnish population) [62,63].

The study protocol was reviewed and accepted by the Academic Ethics Committee of the Tampere region (decision 90/2018). All participants agreed to voluntarily participate in the web-based surveys, and they were informed about the purpose of the study and data processing procedures. Data collection was carried out in collaboration with Norstat, a data solutions provider.

Measures

Drinking

We used the 3-item Alcohol Use Disorders Identification Test (AUDIT-C) to measure risky drinking. AUDIT-C is a widely used screening tool for excessive drinking [66-68]. AUDIT-C is considered to perform almost as well as the full 8-item AUDIT as a screening tool for risky drinking [68]. Three items measure frequency of drinking, units of alcohol per drinking occasion, and the frequency of heavy drinking. Each question has multiple response options, with risk points from 0 to 4, and the scale ranges from 0 to 12. A higher score indicates a higher level of risk drinking. The scale showed good internal consistency among time points (T1: α=.75, T2: α=.73, T3: α=.76, T4: α=.76, and T5: α=.77; see Table 1 for details). Our models used AUDIT-C as a continuous variable. We also report the proportion of excess users, using a cutoff of ≥6 points for men and ≥5 points for women [64].

Table 1. Descriptive statistics of the main study variables.

Outcome variable	T1	T2	T3	T4	T5
Risky drinking, AUDIT-C score (range 0-12), mean (SD)	3.83 (2.50)	3.80 (2.44)	3.75 (2.51)	3.73 (2.47)	3.59 (2.51)
Continuous predictors (range 0-1), mean (SD)					
Psychological distress	0.36 (0.17)	0.34 (0.16)	0.34 (0.15)	0.34 (.15)	0.34 (0.16)
Social media identity bubble	0.45 (0.17)	0.46 (0.17)	0.46 (0.17)	0.47 (.16)	0.46 (0.17)
Categorical predictors (code 0/1), %					
Online dating app use (eg, Tinder)	10.12	9.17	9.88	8.81	9.76
Daily social media communication with colleagues about nonwork topics	10.95	9.40	11.31	9.64	9.52
Social media use several times a day	35.00	34.29	38.69	38.57	39.64
Remote work (≥3 days a week)	5.00	4.88	10.48	22.62	30.71

Psychological Distress

We used the 12-item General Health Questionnaire (GHQ-12) to measure psychological distress [69-72]. All 12 items concerning general mood and psychological strain have 4 answer options, rated on a Likert scale, ranging from very positive (0) to negative (3). The total scale ranged from 0 to 36, where higher scores indicate higher psychological distress. This scale was transformed to a scale of 0 to 1. The internal consistency of the scale was *excellent* at all time points (T1: α=.92, T2: α=.90, T3: α=.90, T4: α=.91, and T5: α=.92).

Social Media Identity Bubble

We measured involvement in social media identity bubbles using the 6-item Identity Bubble Reinforcement Scale (IBRS-6) [44]. Involvement in identity bubbles concerns strong social identification, homophily, and reliance on information from others on social media. Respondents were asked to respond to statements such as "On social media, I prefer interacting with people who share similar interests with me," with response options ranging from 1 (*does not describe me at all*) to 7 (*describes me completely*). The original scale ranged from 6 to 42, but we transformed it to a scale of 0 to 1. The internal consistency of the scale was *good* at all time points (T1: α=.84, T2: α=.83, T3: α=.83, T4: α=.83, and T5: α=.85).

Use of Social Media Platforms

The use of different social media platforms and apps was determined by asking the study participants questions about their general use of social media ("How often do you use the following social media platforms?") and their use of social media for work purposes ("How often do you use the following social media platforms for work purposes?"). Both questions were followed by separate items for different social media apps and internet services, such as Facebook, Instagram, Twitter, YouTube, LinkedIn, Microsoft Teams, and "Dating apps" (eg, Tinder). Answer options included "I don't use it," "less than weekly," "weekly," "daily," and "many times a day." All commonly used apps were first analyzed separately by using dummy variables for weekly users. In the final models, we used a dummy variable for use of online dating apps (0=no use and 1=at least sometimes). Sensitivity analyses were performed for weekly and daily users of online dating apps. The models also included a variable for general social media use. This variable was combined from the most used social media platforms (Facebook, Instagram, Twitter, or YouTube), and the dummy variable was created on the basis of social media use several times a day (0=less often than several times a day and 1=several times a day).

Daily Social Media Communication With Colleagues About Nonwork Topics

We used a single-item measure for daily social media communication with colleagues about nonwork topics: "How often do you use social media to keep in touch with your colleagues or work community regarding nonwork-related matters?" Answer options included "I don't use it," "less than weekly," "weekly," "daily," and "many times a day." A dummy variable was created indicating those communicating with their colleagues informally at least once a day. The measure was previously validated with two Finnish cross-sectional samples [23,24].

Remote Work

We measured remote work with a question on whether respondents were working remotely. Those respondents who indicated that they were working remotely at least 3 days a week were considered remote workers in this study.

Control Variables

Control variables included gender, age, education (MA degree or higher), and occupational area (industrial sector vs others). Gender included options for male, female, and other gender. Very few participants selected *other gender*, and none were among those who had participated in all waves (n=840). Hence, gender was used as a dummy variable (0=male, 56.31%; 1=female, 43.69%). Information on participants' education was obtained via a question including 7 categories (see Multimedia Appendix 1 for details). A dummy variable was created to indicate those who have a master's degree or higher from a university (0=no, 75%; 1=yes, 25%). For occupational field, we used the list of International Standard Industrial Classification of All Economic Activities (see details about the 7 broader categories in Multimedia Appendix 1). We used a dummy variable indicating the industrial area of workers (0=no, 70%; 1=yes, 30%) as a control in our analysis.

We used the 15-item Big Five Inventory for personality [73]. All items had responses ranging from 1 to 7, leading to 5 scales ranging from 3 to 21. These were transformed to a scale of 0 to 1 for the models: openness (mean 0.70, SD 0.16), conscientiousness (mean 0.75, SD 0.15), extroversion (mean 0.64, SD 0.21), agreeableness (mean 0.69, SD 0.14), and neuroticism (mean 0.56, SD 0.17). Internal consistency of the traits varied from acceptable (openness: α=.69; conscientiousness α=.67; agreeableness: α=.56; and neuroticism: α=.71) to good (extroversion: α=.87).

Statistical Analyses

Descriptive results of the study are reported in Table 1 and the text. We provide information about general changes in risky drinking during 2019-2021. The main analyses focus on longitudinal predictors of risky drinking using linear multilevel hybrid models. With hybrid models, it is possible to estimate the within-person effect of time-variant variables, while simultaneously considering the between-person effects. Hybrid models combine the strengths of random-effects and fixed-effects approaches and address their shortcomings [74,75]. We ran the hybrid models with the xthybrid command in Stata (version 16.1; StataCorp) [75].

In our models, all main time-varying variables had both within-person and between-person effects. Within-person effects show how changes in predictors over time are associated with the change in the outcome variable. Between-person variables show group differences between individuals. The models also included several between-person control variables.

We first focused on analyzing within-person and between-person effects of social media app use on risky drinking. We performed analyses on the use of such apps, as well as on weekly and daily

use, when possible. The full model reporting our main findings included only those social media apps that were found relevant during this analysis. Different sensitivity analyses were performed, and we also checked interactions with COVID-19 time points (T3-T5) to determine whether some effects were stronger or different during the pandemic.

Results

Descriptive results showed changes in risky drinking measured were relatively small during the COVID-19 time points examined (T3-T5) compared to the pre-pandemic time points (T1-T2). Drinking decreased slightly, especially at T5 (see Table 1). The proportion of excessive drinkers (AUDIT-C score ≥6 for men and AUDIT-C score ≥5 for women) decreased slightly, from 25.24% (212/840) at T1 to 23.57% (198/840) at T5. Among male participants, no change was reported at all between T1 and T5, with 30.87% (146/473) reporting excessive drinking at both time points. However, excessive drinking among women declined from 17.98% (66/367) at T1 to 14.17% (52/367) at T5.

Our results showed that the most used social media apps were not strongly associated with risky drinking (see Multimedia Appendix 2 for details). Weekly use of Yammer had a small within-person effect (B=0.22; SD 0.11; 95% CI 0.01, 0.46; *P*=.04). Use of online dating apps (eg, Tinder) was strongly associated with risky drinking in all models. In particular, between-person effects were strong for users in general (B=1.52; SD 0.34; 95% CI 0.85, 2.19; *P*<.001), weekly users (B=1.85; SD 0.49; 95% CI 0.88, 2.81; *P*<.001), and daily users (B=2.06; SD 0.80; 95% CI 0.50, 3.62; *P*=.02) in models adjusting for age and gender. Therefore, of all the different social media apps, online dating app use was included in our final model.

Analysis based on the full hybrid model showed that of the within-person predictors, only involvement in social media identity bubbles was significantly associated with risky drinking (B=0.32; *P*=.04; see Table 2 for details). We found robust effects on some between-person variables. Online dating app users were more engaged in risky drinking than nonusers were (B=1.33; *P*<.001). Additionally, those respondents who had daily social media communication with colleagues about nonwork topics drank more than others did (B=0.96; *P*=.005). Women (B=−1.35; *P*<.001) and those with an MA degree or higher from university drank less than others did (B=–0.41; *P*=.02). Older age was associated with higher risky drinking (B=0.02; *P*=.01). Workers in the industrial sector drank more than others did (B=0.40; *P*=.02). Lower conscientiousness (B=−1.41; *P*=.01), lower agreeableness (B=−1.55; *P*=.005), and lower neuroticism (B=−1.17; *P*=.04) were associated with higher drinking. In addition, the full model shows that participants drank less during the COVID-19 era than they previously did (B=–0.13; *P*<.001).

The last part of our analysis aimed at reviewing whether these within-person and between-person effects were stronger during the pandemic. This analysis showed that the within-person effect of Tinder use was observed during the pandemic (B=0.29; SD 0.12; 95% CI 0.05, 0.53; *P*=.02). Other relevant interactions with time were not observed.

Table 2. Multilevel linear hybrid regression model showing within-person and between-person effects on drinking.

	B	Robust SE (B)	95% CI	*P* value
Within-person variables				
Psychological distress	−0.02	0.17	−0.35 to 0.32	.92
Social media identity bubble	0.32	0.15	0.02 to 0.62	.04
Use of a dating app (eg, Tinder)	0.16	0.09	−0.02 to 0.34	.08
Use of social media several times a day	−0.01	0.05	−0.12 to 0.09	.82
Daily social media communication with colleagues about nonwork topics	−0.01	0.06	−0.13 to 0.11	.88
Remote work	−0.11	0.05	−0.21 to 0.00	.04
Between-person variables				
Psychological distress	0.56	0.71	−0.83 to 1.95	.43
Social media identity bubble	−0.28	0.65	−1.54 to 0.99	.67
Use of a dating app (eg, Tinder)	1.33	0.33	0.68 to 1.99	<.001
Use of social media several times a day	0.16	0.20	–0.23 to 0.56	.42
Daily social media communication with colleagues about nonwork topics	0.96	0.34	0.29 to 1.64	.005
Remote work (≥3 days a week)	−0.19	0.33	−0.83 to 0.45	.57
Controls				
Female	−1.35	0.16	−1.66 to −1.03	<.001
Age	0.02	0.01	0.00 to 0.03	.02
MA degree or higher	−0.41	0.17	−0.74 to –0.08	.02
Industrial sector	0.40	0.17	0.05 to 0.74	.02
Openness	0.00	0.52	–1.01 to 1.02	>.99
Conscientiousness	−1.41	0.57	−2.53 to −0.29	.01
Extroversion	0.73	0.45	–0.15 to 1.61	.10
Agreeableness	−1.55	0.55	−2.62 to −0.47	.005
Neuroticism	−1.17	0.56	−2.26 to −0.08	.04

Discussion

Principal Results

This longitudinal, 5-wave study investigated changes in risky drinking in 2019-2021. Longitudinal analysis showed that stronger involvement in social media identity bubbles was associated with higher risky drinking over time. A within-person effect of online dating app use (eg, Tinder) was also found, but only during the COVID-19 pandemic. This finding suggests that unusual circumstances have perhaps led to unusual and risky drinking habits. Online dating app users also drank more than others. Moreover, we found that daily social media communication with colleagues about nonwork topics was associated with risky drinking. Altogether, our results suggest that people engaged in drinking as part of involvement in web-based social bubbles. Furthermore, as dating in bars, clubs, and restaurants was not possible during the COVID-19 pandemic, dating apps such as Tinder were considered an alternative. These dating activities were found to be associated with risky drinking, despite the closure of bars.

Comparison With Prior Work

Previous cross-sectional studies conducted during the COVID-19 pandemic have reported decreased drinking, especially in Europe [10], but increased drinking has also been reported [5-8]. Our results indicating a slight decrease in risky drinking are not surprising, because during spring of 2020 and 2021, there were severe restrictions on bars and restaurants, limiting the possibilities for social drinking. Previous longitudinal studies have shown that drinking less is related to lower access to typical drinking locations [10,12,58].

Our finding on the within-person effect of involvement in social media identity bubbles offers an interesting contribution to the literature. Previous studies have especially considered social media postings as motivational determinants of alcohol use [39] and generally discussed exposure to drinking cues on social media websites [25,42]. Our findings expand this notion by considering the potential effect of engaging with similar-minded individuals. We further noted that individuals communicating with colleagues about nonwork-related issues tended to drink more. As our models controlled for personality, we believe these findings reveal something interesting about the role of social

media in the context of drinking during the COVID-19 pandemic.

We discovered that increased use of dating apps was associated with increased risky drinking during the pandemic. Our longitudinal evidence supports previously noted findings that dating app users drink more than nonusers, and this pattern increases during their dates [34-36]. It is likely that alcohol is consumed during the dates, despite bars being closed. Our study contributes to the growing number of studies on risks and risky behavior related to dating apps [76,77].

Our findings further showed that female gender, younger age, university education, nonindustrial occupational field, conscientiousness, agreeableness, and neuroticism were associated with lower levels of drinking. These results reflect the findings of previous studies [1,58].

We found no longitudinal evidence on the effect of psychological distress on risky drinking during the COVID-19 pandemic. This finding contrasts previous findings on the role of psychological distress and mental well-being in drinking during the pandemic [9,14-18]. One potential reason for this might be that in global comparisons, the COVID-19 situation was manageable in Finland during 2020 and spring 2021. For example, a report by *Der Spiegel* indicated that Finland had coped the best with COVID-19 in comparison with 154 other countries [78]. It is feasible that spring 2020 only presented a major stress for people, reflecting increased drinking among those who were distressed [1]. It is, however, quite early to estimate potential long-lasting effects of the COVID-19 pandemic on drinking levels, as the crisis is still ongoing.

Limitations and Strengths

Our study was limited to Finland and its working population. Hence, some risk drinking groups, including full-time students and retired elderly people, were not part of the study. The results only cover the timeframe of 2019-2021, and it is too early to estimate how the crisis and drinking will further develop. Our results are limited to self-reported information on drinking. Despite these limitations, a major strength of our study is the use of multi-wave, longitudinal data. Data collection started before the COVID-19 crisis, enhancing estimation of the effects of the pandemic on drinking. Our data are particularly strong on social media measures, and the response rate for surveys was high. Our modelling strategy used within-between person hybrid models that provide an advanced way of estimating longitudinal effects of independent variables.

Conclusions

This longitudinal study of workers in Finland showed a decrease in risky drinking during the COVID-19 crisis. The main risk factors for increased risky drinking were found to be related to social media use. Involvement in social media identity bubbles was associated with risky drinking over time. Dating app users were more likely to drink more during the pandemic. Moreover, daily social media communication with colleagues about nonwork topics was associated with higher risky drinking. Our results provide significant evidence of the power of social media during unusual times.

The role of social media in drinking should be considered by social media companies such as Tinder to monitor and regulate alcohol-related postings and advertising on their services and promote more healthy ways of social interaction. This is especially essential during a crisis like the COVID-19 pandemic, when social media services present virtually the only possibilities for forming new social relationships due to lockdowns and restrictions on movement. The connection between social media use, involvement in social media identity bubbles, and increase in alcohol consumption are important findings for organizations and policy makers alike. Education and prevention efforts should consider these risk factors and, possibly, be targeted to at-risk groups.

Acknowledgments

This research has received funding from the Finnish Work Environment Fund (Professional Social Media Use and Work Engagement Among Young Adults Project; project number 118055; principal investigator: Atte Oksanen). Data collection was also partly funded by the Faculty of Social Sciences at the Tampere University, Finland. RO is funded by the Finnish Cultural Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Education and occupational area of the participants at T1 (n=840).
[DOCX File , 15 KB - jmir_v23i12e33125_app1.docx]

Multimedia Appendix 2
Descriptive statistics on use of social media platforms and associations with risky drinking in multilevel linear hybrid regression models.
[DOCX File , 15 KB - jmir_v23i12e33125_app2.docx]

References

1. Oksanen A, Savolainen I, Savela N, Oksa R. Psychological stressors predicting increased drinking during the COVID-19 crisis: a longitudinal national survey study of workers in Finland. Alcohol Alcohol 2021 Apr 29;56(3):299-306 [FREE Full text] [doi: 10.1093/alcalc/agaa124] [Medline: 33160284]
2. Rodriguez LM, Litt DM, Stewart SH. Drinking to cope with the pandemic: the unique associations of COVID-19-related perceived threat and psychological distress to drinking behaviors in American men and women. Addict Behav 2020 Nov;110:106532 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106532] [Medline: 32652385]
3. Callinan S, MacLean S. COVID-19 makes a stronger research focus on home drinking more important than ever. Drug Alcohol Rev 2020 Sep 13;39(6):613-615 [FREE Full text] [doi: 10.1111/dar.13125] [Medline: 32657498]
4. Weerakoon SM, Jetelina KK, Knell G. Longer time spent at home during COVID-19 pandemic is associated with binge drinking among US adults. Am J Drug Alcohol Abuse 2021 Jan 02;47(1):98-106. [doi: 10.1080/00952990.2020.1832508] [Medline: 33280423]
5. Barbosa C, Cowell A, Dowd W. Alcohol consumption in response to the COVID-19 pandemic in the United States. J Addict Med 2021;15(4):341-344. [doi: 10.1097/adm.0000000000000767]
6. Daly M, Robinson E. High-risk drinking in midlife before versus during the COVID-19 crisis: longitudinal evidence from the United Kingdom. Am J Prev Med 2021 Feb;60(2):294-297 [FREE Full text] [doi: 10.1016/j.amepre.2020.09.004] [Medline: 33234355]
7. Jackson SE, Garnett C, Shahab L, Oldham M, Brown J. Association of the COVID-19 lockdown with smoking, drinking and attempts to quit in England: an analysis of 2019-20 data. Addiction 2021 May 26;116(5):1233-1244 [FREE Full text] [doi: 10.1111/add.15295] [Medline: 33089562]
8. Nordeck CD, Riehm KE, Smail EJ, Holingue C, Kane JC, Johnson RM, et al. Changes in drinking days among United States adults during the COVID-19 pandemic. Addiction. Epub ahead of print published online on June 22, 2021 2021 [FREE Full text] [doi: 10.1111/add.15622] [Medline: 34159674]
9. Garnett C, Jackson S, Oldham M, Brown J, Steptoe A, Fancourt D. Factors associated with drinking behaviour during COVID-19 social distancing and lockdown among adults in the UK. Drug Alcohol Depend 2021 Feb 01;219:108461 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108461] [Medline: 33454159]
10. Kilian C, Rehm J, Allebeck P, Braddick F, Gual A, Barták M, European Study Group on Alcohol UseCOVID-19. Alcohol consumption during the COVID-19 pandemic in Europe: a large-scale cross-sectional study in 21 countries. Addiction 2021 Dec 09;116(12):3369-3380. [doi: 10.1111/add.15530] [Medline: 34109685]
11. Valente JY, Sohi I, Garcia-Cerde R, Monteiro MG, Sanchez ZM. What is associated with the increased frequency of heavy episodic drinking during the COVID-19 pandemic? Data from the PAHO regional web-based survey. Drug Alcohol Depend 2021 Apr 01;221:108621. [doi: 10.1016/j.drugalcdep.2021.108621] [Medline: 33636598]
12. Ryerson N, Wilson O, Pena A, Duffy M, Bopp M. What happens when the party moves home? the effect of the COVID-19 pandemic on U.S. college student alcohol consumption as a function of legal drinking status using longitudinal data. Transl Behav Med 2021 Apr 07;11(3):772-774 [FREE Full text] [doi: 10.1093/tbm/ibab006] [Medline: 33598696]
13. White HR, Stevens AK, Hayes K, Jackson KM. Changes in alcohol consumption among college students due to COVID-19: effects of campus closure and residential change. J Stud Alcohol Drugs 2020 Nov;81(6):725-730. [doi: 10.15288/jsad.2020.81.725]
14. Buckner JD, Lewis EM, Abarno CN, Morris PE, Glover NI, Zvolensky MJ. Difficulties with emotion regulation and drinking during the COVID-19 pandemic among undergraduates: the serial mediation of COVID-related distress and drinking to cope with the pandemic. Cogn Behav Ther 2021 Feb 01;50(4):261-275. [doi: 10.1080/16506073.2020.1861084] [Medline: 33522892]
15. McPhee MD, Keough MT, Rundle S, Heath LM, Wardell JD, Hendershot CS. Depression, environmental reward, coping motives and alcohol consumption during the COVID-19 pandemic. Front Psychiatry 2020 Oct 30;11:574676 [FREE Full text] [doi: 10.3389/fpsyt.2020.574676] [Medline: 33192708]
16. Oldham M, Garnett C, Brown J, Kale D, Shahab L, Herbec A. Characterising the patterns of and factors associated with increased alcohol consumption since COVID-19 in a UK sample. Drug Alcohol Rev 2021 Sep 03;40(6):890-899 [FREE Full text] [doi: 10.1111/dar.13256] [Medline: 33657644]
17. Sallie SN, Ritou V, Bowden-Jones H, Voon V. Assessing international alcohol consumption patterns during isolation from the COVID-19 pandemic using an online survey: highlighting negative emotionality mechanisms. BMJ Open 2020 Nov 26;10(11):e044276 [FREE Full text] [doi: 10.1136/bmjopen-2020-044276] [Medline: 33243820]
18. Tran TD, Hammarberg K, Kirkman M, Nguyen HTM, Fisher J. Alcohol use and mental health status during the first months of COVID-19 pandemic in Australia. J Affect Disord 2020 Dec 01;277:810-813 [FREE Full text] [doi: 10.1016/j.jad.2020.09.012] [Medline: 33065821]
19. Boyd D. Social media: a phenomenon to be analyzed. Social Media Soc 2015 May 11;1(1):205630511558014. [doi: 10.1177/2056305115580148]
20. Boyd D, Ellison N. Social network sites: definition, history, and scholarship. J Comput Mediat Commun 2007;13:210-230. [doi: 10.1111/j.1083-6101.2007.00393.x]
21. Keipi T, Näsi M, Oksanen A, Räsänen P. Online Hate and Harmful Content: Cross-National Perspectives. New York, NY: Routledge; 2017.

22. McCay-Peet L, Quan-Haase A. What is social mediawhat questions can social media research help help us answer. In: The SAGE handbook of social media research methods. London: SAGE Publications; 2017.
23. Oksa R, Kaakinen M, Savela N, Ellonen N, Oksanen A. Professional social media usage: Work engagement perspective. New Media & Society 2020 May 22;23(8):2303-2326. [doi: 10.1177/1461444820921938]
24. Oksa R, Saari T, Kaakinen M, Oksanen A. The motivations for and well-being implications of social media use at work among millennials and members of former generations. Int J Environ Res Public Health 2021 Jan 19;18(2):803 [FREE Full text] [doi: 10.3390/ijerph18020803] [Medline: 33477818]
25. Moewaka Barnes H, McCreanor T, Goodwin I, Lyons A, Griffin C, Hutton F. Alcohol and social media: drinking and drunkenness while online. Critical Public Health 2015 Jun 26;26(1):62-76. [doi: 10.1080/09581596.2015.1058921]
26. Westgate EC, Holliday J. Identity, influence, and intervention: the roles of social media in alcohol use. Curr Opin Psychol 2016 Jun;9:27-32. [doi: 10.1016/j.copsyc.2015.10.014]
27. Boyle SC, LaBrie JW, Froidevaux NM, Witkovic YD. Different digital paths to the keg? How exposure to peers' alcohol-related social media content influences drinking among male and female first-year college students. Addict Behav 2016 Jun;57:21-29 [FREE Full text] [doi: 10.1016/j.addbeh.2016.01.011] [Medline: 26835604]
28. Brunborg GS, Burdzovic Andreas J. Increase in time spent on social media is associated with modest increase in depression, conduct problems, and episodic heavy drinking. J Adolesc 2019 Jul;74:201-209. [doi: 10.1016/j.adolescence.2019.06.013] [Medline: 31254779]
29. Brunborg GS, Andreas JB, Kvaavik E. Social media use and episodic heavy drinking among adolescents. Psychol Rep 2017 Jun 06;120(3):475-490. [doi: 10.1177/0033294117697090] [Medline: 28558617]
30. LaBrie JW, Boyle SC, Young SH, Tan CN. Prospective relationships between objectively assessed social media use, drinking norms, and alcohol consumption among first-year students. J Stud Alcohol Drugs 2021 May;82(3):339-350. [doi: 10.15288/jsad.2021.82.339]
31. Ng Fat L, Cable N, Kelly Y. Associations between social media usage and alcohol use among youths and young adults: findings from Understanding Society. Addiction 2021 Nov 22;116(11):2995-3005. [doi: 10.1111/add.15482] [Medline: 33886119]
32. Ridout B. Facebook, social media and its application to problem drinking among college students. Curr Opin Psychol 2016 Jun;9:83-87. [doi: 10.1016/j.copsyc.2015.12.005]
33. Savolainen I, Oksanen A, Kaakinen M, Sirola A, Miller BL, Paek HJ, et al. The association between social media use and hazardous alcohol consumption among youths: a four-country study. Alcohol Alcohol 2020 Feb 07;55(1):86-95. [doi: 10.1093/alcalc/agz088] [Medline: 31761930]
34. Erevik EK, Kristensen JH, Torsheim T, Vedaa Ø, Pallesen S. Tinder use and romantic relationship formations: a large-scale longitudinal study. Front Psychol 2020 Aug 14;11:1757 [FREE Full text] [doi: 10.3389/fpsyg.2020.01757] [Medline: 32922327]
35. Flesia L, Fietta V, Foresta C, Monaro M. The association between dating apps and alcohol consumption in an Italian sample of active users, former users, and non-users. Social Sciences 2021 Jul 01;10(7):249. [doi: 10.3390/socsci10070249]
36. Fansher AK, Eckinger S. Tinder tales: an exploratory study of online dating users and their most interesting stories. Deviant Behav 2020 Mar 04;42(9):1194-1208. [doi: 10.1080/01639625.2020.1734170]
37. Rogge RD, Crasta D, Legate N. Is Tinder-Grindr use rRisky? distinguishing venue from individuals' behavior as unique predictors of sexual risk. Arch Sex Behav 2020 May 12;49(4):1263-1277. [doi: 10.1007/s10508-019-01594-w] [Medline: 31832853]
38. Tomaszewska P, Schuster I. Comparing sexuality-related cognitions, sexual behavior, and acceptance of sexual coercion in dating app users and non-users. Sex Res Soc Policy 2019 Jun 22;17(2):188-198. [doi: 10.1007/s13178-019-00397-x]
39. Erevik EK, Torsheim T, Andreassen CS, Vedaa Ø, Pallesen S. Disclosure and exposure of alcohol on social media and later alcohol use: a large-scale longitudinal study. Front Psychol 2017 Nov 01;8:1934 [FREE Full text] [doi: 10.3389/fpsyg.2017.01934] [Medline: 29163308]
40. Gupta H, Lam T, Pettigrew S, Tait RJ. The association between exposure to social media alcohol marketing and youth alcohol use behaviors in India and Australia. BMC Public Health 2018 Jun 13;18(1):726 [FREE Full text] [doi: 10.1186/s12889-018-5645-9] [Medline: 29895264]
41. Hendriks H, Van den Putte B, Gebhardt WA, Moreno MA. Social drinking on social media: content analysis of the social aspects of alcohol-related posts on Facebook and Instagram. J Med Internet Res 2018 Jun 22;20(6):e226 [FREE Full text] [doi: 10.2196/jmir.9355] [Medline: 29934290]
42. Curtis BL, Lookatch SJ, Ramo DE, McKay JR, Feinn RS, Kranzler HR. Meta-analysis of the association of alcohol-related social media use with alcohol consumption and alcohol-related problems in adolescents and young adults. Alcohol Clin Exp Res 2018 Jun 22;42(6):978-986 [FREE Full text] [doi: 10.1111/acer.13642] [Medline: 29786874]
43. Litt DM, Rodriguez LM, Stewart SH. Examining associations between social networking site alcohol-specific social norms, posting behavior, and drinking to cope during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw. Epub ahead of print posted online on March 16, 2021 2021. [doi: 10.1089/cyber.2020.0568] [Medline: 33733832]

44. Kaakinen M, Sirola A, Savolainen I, Oksanen A. Shared identity and shared information in social media: development and validation of the identity bubble reinforcement scale. Media Psychology 2018 Nov 17;23(1):25-51. [doi: 10.1080/15213269.2018.1544910]
45. Koivula A, Kaakinen M, Oksanen A, Räsänen P. The role of political activity in the formation of online identity bubbles. Policy & Internet 2019 Jul 15;11(4):396-417. [doi: 10.1002/poi3.211]
46. Coronavirus disease (COVID-19): Small public gatherings. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-small-public-gatherings# [accessed 2021-08-06]
47. Living, working and COVID-19 (Update April 2021): mental health and trust decline across EU as pandemic enters another year. Eurofound. URL: https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef21064en.pdf [accessed 2021-08-06]
48. Living, Working and COVID-19. COVID-19 Series. Eurofound. URL: https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef20059en.pdf [accessed 2021-08-06]
49. Kestilä L, Härmä V, Rissanen V. COVID-19-epidemian vaikutukset hyvinvointiin, palvelujärjestelmään ja kansantalouteen [The implications of COVID-19 to Wellbeing, Service Systems and Economy]. Helsinki: Finnish Institute of Health and Welfare; 2020.
50. 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. Lancet Psychiatry 2020 Apr;7(4):e15-e16. [doi: 10.1016/s2215-0366(20)30078-x]
51. Lorant V, Smith P, Van den Broeck K, Nicaise P. Psychological distress associated with the COVID-19 pandemic and suppression measures during the first wave in Belgium. BMC Psychiatry 2021 Feb 18;21(1):112 [FREE Full text] [doi: 10.1186/s12888-021-03109-1] [Medline: 33602149]
52. Choi N, DiNitto D. Psychological distress, binge/heavy drinking, and gender differences among older adults. Am J Addict 2011;20(5):420-428. [doi: 10.1111/j.1521-0391.2011.00149.x] [Medline: 21838840]
53. Rehm J, Kilian C, Ferreira-Borges C, Jernigan D, Monteiro M, Parry CDH, et al. Alcohol use in times of the COVID 19: implications for monitoring and policy. Drug Alcohol Rev 2020 May;39(4):301-304 [FREE Full text] [doi: 10.1111/dar.13074] [Medline: 32358884]
54. Adan A, Forero DA, Navarro JF. Personality traits related to binge drinking: a systematic review. Front Psychiatry 2017 Jul 28;8:134 [FREE Full text] [doi: 10.3389/fpsyt.2017.00134] [Medline: 28804465]
55. Foster J, Canfield M. Predictors of hazardous drinking among home drinkers. J Subst Use 2017 Mar 28;22(6):637-642. [doi: 10.1080/14659891.2017.1296040]
56. Global status report on alcohol and health 2018. World Health Organization. 2018 Sep 27. URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2021-11-12]
57. Bollen Z, Pabst A, Creupelandt C, Fontesse S, Lannoy S, Pinon N, et al. Prior drinking motives predict alcohol consumption during the COVID-19 lockdown: A cross-sectional online survey among Belgian college students. Addict Behav 2021 Apr;115:106772. [doi: 10.1016/j.addbeh.2020.106772] [Medline: 33418433]
58. Bramness JG, Bye EK, Moan IS, Rossow I. Alcohol use during the COVID-19 pandemic: self-reported changes and motives for change. Eur Addict Res 2021 Apr 9;27(4):257-262 [FREE Full text] [doi: 10.1159/000515102] [Medline: 33839730]
59. Wardell JD, Kempe T, Rapinda KK, Single A, Bilevicius E, Frohlich JR, et al. Drinking to cope during COVID-19 pandemic: the role of external and internal factors in coping motive pathways to alcohol use, solitary drinking, and alcohol problems. Alcohol Clin Exp Res 2020 Oct;44(10):2073-2083. [doi: 10.1111/acer.14425] [Medline: 32870516]
60. Oksanen A, Kaakinen M, Latikka R, Savolainen I, Savela N, Koivula A. Regulation and trust: 3-month follow-up study on COVID-19 mortality in 25 European countries. JMIR Public Health Surveill 2020 Apr 24;6(2):e19218 [FREE Full text] [doi: 10.2196/19218] [Medline: 32301734]
61. Oksanen A, Oksa R, Kaakinen M, Savela N, Latikka R, Ellonen N. Sosiaalinen media työelämässä: hyvinvoinnin ulottuvuuksia. Book in Finnish. Tampere: Tampere University: Emerging Technologies Lab; 2020.
62. Statistics Finland. Employment. Helsinki: Statistics Finland; 2021. URL: http://www.stat.fi/til/tyokay/index_en.html [accessed 2021-11-14]
63. Labour force survey e-publication. Statistics Finland. URL: http://www.stat.fi/til/tyti/index_en.html [accessed 2021-11-12]
64. Yearbook of Alcohol and Drug Statistics 2020 [Book in Finnish]. In: Official Statistics of Finland (OSF). Helsinki: Statistics Finland; 2021.
65. Use of the internet for following the media and for communication has increased. Statistics Finland. 2020 Nov 10. URL: http://www.stat.fi/til/sutivi/2020/sutivi_2020_2020-11-10_tie_001_en.html [accessed 2021-11-12]
66. Babor T, Higgins-Biddle J, Saunders J. AUDIT, The alcohol use disorders identification test: Guidelines for use in primary care. Geneva: WHO; 2001.
67. Higgins-Biddle JC, Babor TF. A review of the Alcohol Use Disorders Identification Test (AUDIT), AUDIT-C, and USAUDIT for screening in the United States: Past issues and future directions. Am J Drug Alcohol Abuse 2018 May 03;44(6):578-586 [FREE Full text] [doi: 10.1080/00952990.2018.1456545] [Medline: 29723083]

68. Kaarne T, Aalto M, Kuokkanen M, Seppä K. AUDIT-C, AUDIT-3 and AUDIT-QF in screening risky drinking among Finnish occupational health-care patients. Drug Alcohol Rev 2010 Sep;29(5):563-567. [doi: 10.1111/j.1465-3362.2010.00172.x] [Medline: 20887582]
69. Banks M, Clegg C, Jackson P, Kemp N, Stafford E, Wall T. The use of the General Health Questionnaire as an indicator of mental health in occupational studies. J Occup Psychol 1980 Sep;3:94. [doi: 10.1111/j.2044-8325.1980.tb00024.x]
70. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med 1979 Feb;9(1):139-145. [doi: 10.1017/s0033291700021644] [Medline: 424481]
71. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan 01;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
72. Kalliath TJ, O'Driscoll MP, Brough P. A confirmatory factor analysis of the General Health Questionnaire-12. Stress and Health 2004 Jan 29;20(1):11-20. [doi: 10.1002/smi.993]
73. Hahn E, Gottschling J, Spinath FM. Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S). J Res Pers 2012 Jun;46(3):355-359. [doi: 10.1016/j.jrp.2012.03.008]
74. Schunck R. Within and between estimates in random-effects models: advantages and drawbacks of correlated random effects and hybrid models. Stata J 2013 Mar 01;13(1):65-76. [doi: 10.1177/1536867x1301300105]
75. Schunck R, Perales F. Within- and between-cluster effects in generalized linear mixed models: a discussion of approaches and the xthybrid command. Stata J 2017 Mar 01;17(1):89-115. [doi: 10.1177/1536867x1701700106]
76. Castro Á, Barrada JR. Dating apps and their sociodemographic and psychosocial correlates: a systematic review. Int J Environ Res Public Health 2020 Sep 07;17(18):6500 [FREE Full text] [doi: 10.3390/ijerph17186500] [Medline: 32906632]
77. Kaakinen M, Koivula A, Savolainen I, Sirola A, Mikkola M, Zych I, et al. Online dating applications and risk of youth victimization: a lifestyle exposure perspective. Aggress Behav 2021 Sep 26;47(5):530-543. [doi: 10.1002/ab.21968] [Medline: 34038595]
78. Dambeck H, Hoffman C. Wo der Kampf gegen die Pandemie gelingt. Article in German. Spiegel Ausland. 2021 Jul 08. URL: https://www.spiegel.de/ausland/corona-laendervergleich-wo-der-kampf-gegen-die-pandemie-gelingt-a-b253bbd9-fe9b-4d5e-9169-63b57d771216 [accessed 2021-08-06]

Abbreviations

AUDIT-C: 3-item Alcohol Use Disorders Identification Test
GHQ-12: General Health Questionnaire
IBRS: 6-item Identity Bubble Reinforcement Scale

Edited by C Basch; submitted 24.08.21; peer-reviewed by M Barratt, T Tillmann, N Nazari; comments to author 10.10.21; revised version received 12.10.21; accepted 13.10.21; published 02.12.21.

Please cite as:
Oksanen A, Oksa R, Savela N, Celuch M, Savolainen I
Drinking and Social Media Use Among Workers During COVID-19 Pandemic Restrictions: Five-Wave Longitudinal Study
J Med Internet Res 2021;23(12):e33125
URL: https://www.jmir.org/2021/12/e33125
doi:10.2196/33125
PMID:34662290

Original Paper

User Behaviors and User-Generated Content in Chinese Online Health Communities: Comparative Study

Yuqi Lei[1*], BS; Songhua Xu[1*], PhD; Linyun Zhou[1], MPH

Institute of Medical Artificial Intelligence, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Songhua Xu, PhD
Institute of Medical Artificial Intelligence
The Second Affiliated Hospital of Xi'an Jiaotong University
No 5 Jianqiang Road
Xincheng District
Xi'an, 710016
China
Phone: 86 029 82666758
Email: 36832164@qq.com

Abstract

Background: Online health communities (OHCs) have increasingly gained traction with patients, caregivers, and supporters globally. Chinese OHCs are no exception. However, user-generated content (UGC) and the associated user behaviors in Chinese OHCs are largely underexplored and rarely analyzed systematically, forfeiting valuable opportunities for optimizing treatment design and care delivery with insights gained from OHCs.

Objective: This study aimed to reveal both the shared and distinct characteristics of 2 popular OHCs in China by systematically and comprehensively analyzing their UGC and the associated user behaviors.

Methods: We concentrated on studying the lung cancer forum (LCF) and breast cancer forum (BCF) on Mijian, and the diabetes consultation forum (DCF) on Sweet Home, because of the importance of the 3 diseases among Chinese patients and their prevalence on Chinese OHCs in general. Our analysis explored the key user activities, small-world effect, and scale-free characteristics of each social network. We examined the UGC of these forums comprehensively and adopted the weighted knowledge network technique to discover salient topics and latent relations among these topics on each forum. Finally, we discussed the public health implications of our analysis findings.

Results: Our analysis showed that the number of reads per thread on each forum followed gamma distribution (H_L=0, H_B=0, and H_D=0); the number of replies on each forum followed exponential distribution (adjusted R_L^2=0.946, adjusted R_B^2=0.958, and adjusted R_D^2=0.971); and the number of threads a user is involved with (adjusted R_L^2=0.978, adjusted R_B^2=0.964, and adjusted R_D^2=0.970), the number of followers of a user (adjusted R_L^2=0.989, adjusted R_B^2=0.962, and adjusted R_D^2=0.990), and a user's degrees (adjusted R_L^2=0.997, adjusted R_B^2=0.994, and adjusted R_D^2=0.968) all followed power-law distribution. The study further revealed that users are generally more active during weekdays, as commonly witnessed in all 3 forums. In particular, the LCF and DCF exhibited high temporal similarity (ρ=0.927; P<.001) in terms of the relative thread posting frequencies during each hour of the day. Besides, the study showed that all 3 forums exhibited the small-world effect (mean σ_L=517.15, mean σ_B=275.23, and mean σ_D=525.18) and scale-free characteristics, while the global clustering coefficients were lower than those of counterpart international OHCs. The study also discovered several hot topics commonly shared among the 3 disease forums, such as disease treatment, disease examination, and diagnosis. In particular, the study found that after the outbreak of COVID-19, users on the LCF and BCF were much more likely to bring up COVID-19–related issues while discussing their medical issues.

Conclusions: UGC and related online user behaviors in Chinese OHCs can be leveraged as important sources of information to gain insights regarding individual and population health conditions. Effective and timely mining and utilization of such content can continuously provide valuable firsthand clues for enhancing the situational awareness of health providers and policymakers.

(J Med Internet Res 2021;23(12):e19183) doi:10.2196/19183

KEYWORDS

online health community; user behaviors; user-generated content; social network analysis; weighted knowledge network

Introduction

Background

An online community is a social group created by internet users for a variety of purposes and interests. The rapid development of the "internet plus" [1] technology has further promoted the value of online communities in recent years. In the meanwhile, the health consciousness of populations and their motivation for better self-health management have been steadily growing. Propelled by the strong desire to mitigate the information asymmetry between doctors and patients pervasive in traditional health care communications, patients have gained a new way to share their disease situation and receive needed health care advice through online health communities (OHCs). For example, a continuously rising number of patients and their relatives continually participate in OHCs, actively share their treatment experiences, and openly express their personal opinions and feelings on various issues encountered during treatment or the whole care journey. The value of OHCs for exchanging emotional communications and delivering social support for patients and their families has also been widely recognized [2]. At present, an increasing multitude of user-generated content (UGC) and associated online user behaviors are becoming available on OHCs. The US Office of the National Coordinator for Health Information Technology defines patient-generated health data as health-related data created, recorded, or gathered by or from patients (or family members or other caregivers) to help address a health concern [3]. The Chinese National Health Commission publicly released an Action Plan for the Further Improvement of Medical Services (2018-2020), which emphasizes the role of patient organizations in knowledge sharing, whole-course disease management, rehabilitation support, drug development, and clinical trials [4]. However, such UGC in Chinese OHCs and their associated user behaviors are often underexplored and rarely analyzed systematically, thus losing valuable clues and evidence for improving treatment design and patient care.

Status of Research Concerning OHCs

A number of OHCs exist, providing users with diverse and fluent ways to exchange information, share experiences, seek answers, and receive support. PatientsLikeMe is the first and also the largest social network platform in the world dedicated to patients. By 2018, more than 650,000 users had communicated and shared their health information over the platform, with more than 2900 diseases involved [5]. MyHealthTeams [6] is a social network for people living with chronic diseases, which aims to provide mutual aids for its participants, and has gathered more than 2 million users spread over 33 online disease platforms. Several OHCs, such as Breastcancer [7] and BecomeAnEX [8], are also popular among patients. In China, Manyoubang [9] provides an interactive OHC for patients with chronic diseases, which has more than 22 subforums. Yi Xiang Network [10] is the largest case sharing website in China, providing services, such as case upload, communications, and mutual assistance, for patients. Mijian [11] is the largest interactive OHC for patients in China at the time of writing this manuscript, which integrates multiple single disease forums with interactive question and answer functions. Other OHCs in China, such as Sweet Home [12] and Lymphoma Home [13], mainly focus on servicing patients with chronic conditions. Meanwhile, in some general-purpose Chinese online forums (eg, Tianya [14], Tieba [15], and Zhihu [16]), there are also subforums specifically dedicated to disease-centric discussions.

Since the emergence of versatile OHCs, scholars have attempted to analyze these virtual communities from various perspectives. For example, Smailhodzic et al [17] conducted a literature review covering 22 articles, according to which, patients' use of social media were classified into the following 6 categories: emotional, information, esteem, network support, social comparison, and emotional expression. Dongxiang [18] overviewed OHCs in China from 3 perspectives, including their UGC, and the characteristics of participants and underlying online communities. Wu et al [19] summarized research hotspots concerning OHCs and the evolution of OHCs, as well as the key analysis methods for OHCs. To reveal factors that may motivate knowledge sharing in OHCs, scholars utilized text mining to better understand and predict user participation [20,21]. Fernandes et al [22] adopted a netnography method to analyze the positive impact of OHCs on the prognosis of diabetes. Li [23] utilized the structural equation model to study factors affecting individual patient's willingness to share medical information. Scholars also studied the distribution of health topics according to questions and answers on OHCs using machine learning approaches [24-27]. In addition, scholars utilized social network analysis methods to analyze knowledge exchange behaviors among users in OHCs by constructing and examining underlying knowledge-sharing networks [28-33].

Overall, existing research on OHCs has mainly focused on uncovering users' motivation for participating in the OHCs, discussing factors affecting users' online knowledge-sharing behaviors, and mining UGC in OHCs. For Chinese OHCs, existing research primarily concentrates on examining small-scale single-disease forums. In comparison with peer international studies, both the breadth and depth of current analysis regarding Chinese OHCs are much more limited, calling for expanded efforts to broaden the understanding and strengthen preliminary findings produced through existing studies. To meet the demand and fill the gap, this study comprehensively examined 3 representative disease forums hosted on the 2 most popular Chinese OHCs. The large-scale evaluation reveals both the shared traits and distinct characteristics of user behaviors and UGC on these forums to shed light on understanding user behaviors and UGC on Chinese OHCs in general.

Objectives

Given the popularity and proliferation of OHCs, understanding multifaceted patient experiences reflected from UGC in these forums and related user behaviors can provide many valuable

insights for enhancing public health awareness and improving the quality of the care delivered. Comprehensive and in-depth analysis of such user content and behaviors can also help optimize the design and management of OHCs from a software engineering perspective, as well as the design and development of better community-based knowledge services at large. Driven by the above anticipated benefits, this study performed an in-depth analysis on UGC and related online user behaviors in 3 large-scale OHCs in China. We utilized a variety of social network analysis methods and constructed a knowledge-sharing network for each OHC to study the evolution of OHCs, discover characteristics of user behaviors, uncover salient topics and their relations in each of the virtual communities, and reveal common traits and distinct characteristics in the 3 OHCs examined. Through these case analyses, we also aimed to offer insights regarding user behaviors and UGC in Chinese OHCs in general.

Methods

Data Collection

Two influential OHCs in China, Mijian and Sweet Home, were selected for analysis in this study. Mijian was selected for analysis in this study because it is the largest OHC for patients in China at present. The site targets to serve patients diagnosed with chronic, severe, or rare diseases, aiming to help relieve their psychological stresses, learn disease-related health knowledge, and effectively acquire medical resources. Sweet Home was selected for analysis in this study because it is the largest OHC in China for patients with diabetes. The site offers categorized forums for medical consultation, service guidance, and emotional expression. Through the site, patients with diabetes can not only discuss their medical conditions, but also connect and communicate remotely with other patients across the country. Regarding the 2 focus OHCs identified in this study, our analysis concentrated on examining the lung cancer forum (LCF) and breast cancer forum (BCF) on Mijian, and the diabetes consultation forum (DCF) on Sweet Home in particular because of their predominant popularity among users of the 2 OHCs and the significance of the 3 diseases for the well-being of Chinese patients and the population as a whole given that breast cancer and lung cancer represent the 2 leading chronic noncommunicable diseases in the world and China has the largest number of patients with diabetes globally [34,35].

In a disease forum, a single conversation is referred to as a "thread" (ie, a topic). Users can respond to another person's thread, which is referred to as a "reply." Thus, a post made by a user on a forum can either be an original thread created by the user or a reply to another user's thread [36]. We crawled all threads on the 3 focus disease forums from their respective forum establishment dates (LCF: November 15, 2013; BCF: August 25, 2015; DCF: September 1, 2005) to October 20, 2020. Once a post was crawled, we also obtained its ID, posting time, number of reads, and number of replies. We then performed a series of data cleaning operations, including filtering posts automatically created by chatbots on these forums and deleting missing data. Table 1 presents key statistics of the acquired online data sets in comparison with those of data sets used in peer studies [20,24,26,28,30,37,38], which shows that the scale of the current analysis significantly transcends that of all prior efforts.

Table 1. Comparison between experimental data sets analyzed in this study and the counterpart data sets in peer studies.

Study	Website	Forum	Number of threads	Number of users	Number of replies
Present study	Mijian [11]	Lung cancer forum	37,090	22,610	254,687
Present study	Mijian [11]	Breast cancer forum	112,790	31,909	2,123,728
Present study	Sweet Home [12]	Diabetes consultation forum	41,060	26,751	466,225
Wu et al [28][a]	Yi Xiang Network [13]	Breast cancer forum	754	540	3498
Wu et al [37][a]	39 Health Network [39]	Hepatitis B forum	1066	N/A[c]	N/A
Wu et al [38][a]	Tieba [15]	Tumor forum	2009	1476	11,940
Shi et al [30][a]	Manyoubang [9]	Diabetes mutual aid forum	777	636	3553
Wang et al [20][b]	Breastcancer [7]	Breast cancer forum	107,549	49,552	2,800,000
Wang et al [24][b]	BecomeAnEX [8]	Smoking cessation	38,156	5435	316,886
Della Rosa et al [26][b]	Facebook [40]	Multiple sclerosis	N/A	24,915	N/A

[a]Online health communities in China.

[b]Online health communities in other countries.

[c]N/A: not applicable.

Social Network Analysis

A social network is a social structure made up of a set of social actors (such as individuals and organizations), sets of dyadic ties, and other social interactions between actors [41]. Social network analysis can help identify community structures at the network level, as well as individual behaviors at the single-user level. Since a user could be both a thread author and a reply author, this study adopted a directed network structure to model the community network. In such a directed network, each edge

of the network is directional, where the in-degree of a node refers to the number of directed edges ending with the node. Conversely, the out-degree of a node is the number of directed edges starting from the node. The total degree of a node is the total number of its network neighbors irrespective of the tie direction (ie, the sum of its in-degree and out-degree).

We conducted a topological analysis for complex networks [41] to explore the structural characteristics of each forum. Our analysis was carried out in 2 steps. First, we explored the small-world effect of each social network. We took the path length between 2 nodes as the minimum number of edges connecting these nodes in the network. The average path length, also known as characteristic path length, is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. Let d_{ij} denote the shortest distance between 2 nodes i and j in the network. Assume that $d_{ij}=0$ if $i=j$ or j cannot be reached from i. Then, the average path length L is as follows:

where N is the number of network nodes. The clustering coefficient of a network measures the degree of node clustering in the network. Assume a node k has n adjacent neighboring nodes (N_1, N_2, …, N_n). If 2 nodes i and j are connected with a link, the directed link is denoted as e_{ij}. The local clustering coefficient of the node k is defined as follows:

Assume the entire network has K nodes in total. The average clustering coefficient of the network is the mean of the clustering coefficients of all its nodes, that is,

The small-world effect, also known as the 6 degrees of separation, is the idea that all strangers can be related through 6 or fewer people [42]. Watts et al proposed a small-world network model (WS model) [43], in which a small-world network is characterized by a small average path length and a high clustering coefficient. As a general method for quantifying the small-world effect of a network, the network can be measured by comparing its clustering coefficient and average path length with those of an equivalent Erdös–Rényi (ER) random network that has the same number of nodes and edges [44]. To construct such an equivalent random network, we employed the following generative procedure: Let N and M be the number of nodes and edges expected of the network to be generated, respectively. The network is initialized to have N unconnected nodes. At each step, we randomly selected and linked a pair of nodes not currently connected in the network. We repeated the above step until all M edges were added into the network. Given the random network generated, σ can be calculated as follows:

where  is the average clustering coefficient of the network, L is the average path length of the network, C_r is the average clustering coefficient of the equivalent ER random network, and L_r is the average path length of the equivalent ER random network. If $\sigma>1$ (eg, ), the network is considered a small-world network [44], in which case, it is assumed that knowledge can be spread efficiently and rapidly in the community represented by the network.

Second, we explored the scale-free property of social networks. Scale-free property is a structural characteristic concerning a network as introduced by Barabási et al [45]. Across scientific domains and classes of networks, it is common to encounter the claim that most or all real-world networks are scale free. Generally, a network is deemed scale free if the fraction of nodes with degree k follows a power-law distribution $P(k)=c\times k^{-r}$, where $r>1$. The property mainly comprises the following 2 aspects. First, the distribution of nodes follows a power-law distribution, where most nodes have few links, while a small fraction of nodes has a large number of links. In a power-law distribution, the probability that a node has degree k follows the distribution equation $P(k)=c\times k^{-r}$, where c and r are network-specific constants. It is generally believed that if the degree of nodes in a network follows a power-law distribution, the network is a scale-free network. Second, during a network growth process, new nodes preferentially establish relations with well-connected nodes. The scale-free network model is also commonly referred to as the B-A model [45].

Weighted Knowledge Network

Topic analysis techniques can be leveraged to extract conceptual topics, determine their types, and analyze their internal structures latent in a large text corpus. In this study, we analyzed health topics on OHCs to identify hot topics and the salient health information needs of their users.

We executed the topic analysis in 2 steps. First, we extracted key phrases among UGC according to point-wise mutual information (PMI), as well as the left and right information entropy, with which the co-occurrence relationship between words can be efficiently found. In this step, mutual information is mainly used to measure the degree of correlation between 2 signals according to information theory [46], which is repurposed to measure the degree of interdependence between 2 variables. In natural language processing (NLP), PMI is used to calculate the degree of correlation between 2 words, so that the co-occurrence of words can be found from a statistical perspective to examine whether any semantic correlation or thematic correlation exists between a pair of words. The PMI of 2 adjacent words x and y is computed as follows:

where $p(x)$ is the probability of word x appearing in all threads, that is, $p(x)$=the number of occurrences of word x / the total number of words in all threads; $p(y)$ is the probability of word y appearing in all threads, that is, $p(y)$=the number of occurrences of word y / the total number of words in all threads; and $p(x,y)$ is the joint probability of x and y, which is the

probability that 2 words (x,y) appear adjacent to each other in the text. A higher PMI of x and y is associated with higher internal aggregation and greater possibility of the 2 words forming a phrase. Conversely, those 2 words are more likely to have phrasal boundaries.

Entropy is an uncertainty measure associated with a random variable. A higher entropy is associated with greater underlying information content and hence higher uncertainty [47]. In NLP, the left and right entropy of the word W are defined as follows:

where E_L and E_R are the left entropy and right entropy of the word W, respectively; A and B represent the sets of all words appearing to the left and right of W, respectively; and a and b represent the words appearing immediately on the left and right sides of W, respectively. Greater left or right entropy is associated with a higher degree of freedom of the word, which indicates more abundant choices for a target word surrounding the given word W.

Second, we treated a keyword as a node and the co-occurrence relationship between a pair of keywords as an edge to construct a weighted knowledge network (WKN) [37,48]. In the process, we also assigned weights to the nodes and edges according to weights of the corresponding keywords and the relationship strength between the corresponding key phrases. The WKN integrated and modeled fragmentation knowledge of the thematic content, which can be used to effectively discover the internal relationships and overall characteristics of a knowledge network. More specifically, we summed the PMI value of the left entropy and right entropy calculated above, which was used as a measure of 2 words as a phrase. We then extracted key phrases in each post and their respective weights where the frequency of a key phrase is used as the weight of the key phrase. We define E as the keywords co-occurrence set and $Q(E)$ as the weight set associated with E, as follows:

In the above equations, if keywords k_i and k_j formulate a key phrase, as indicated by a co-occurrence relationship between them in the WKN, then e_{ij}=1; otherwise, e_{ij}=0. $q(e_{ij})$ is the weight of e_{ij}. $q(e_{ij})=n(e_{ij})/N$, where $n(e_{ij})$ is the number of occurrences of a key phrase in all phrases and N is the total number of all phrases. Next, all detected key phrases were organized as a keywords set K, for which an associated keyword weight set $Q(K)$ was introduced as follows:

where k_i is a keyword and $q(k_i)$ is its weight. $q(k_i)=m(k_i)/M$, where $m(k_i)$ is the number of occurrences of a keyword k_i in the 200 key phrases and M is the total number of all keywords in the 200 key phrases.

Now, we can define a WKN model for the concerned OHC as follows:

According to the constructed WKN model, the results can be displayed by social network visualization tools.

Results

Descriptive Statistics

Table 2 presents descriptive statistics of the OHC data analyzed in this study where extreme outliers were removed during the preprocessing. In terms of the number of reads and replies per thread, the data distribution was severely nonuniform. In other words, most of the threads had fewer reads and replies, and only a few threads got a large number of reads and replies, which meant that most users preferentially read the threads that had received a larger number of replies, thus resulting in the polarization. The SD is a measure of the amount of variation or dispersion of a set of numbers, which is also affected by the volume of data analyzed. The coefficient of variation is a statistical measure of the dispersion of data points around the mean of a data series. The coefficient of variation of the general normal distribution is less than 1. Considering the large coefficient of variation for each attribute listed in Table 2 and according to the skewness of the frequency distribution graph (SK_L=35.41, SK_B=25.65, and SK_D=12.45), we concluded that none of them follows a normal distribution.

Table 2. Statistical characteristics of the 3 data sets analyzed in this study.

Data set and variable	Minimum	Q1	Median	Q3	Maximum	Mean	SD	CV[a]
Lung cancer forum								
Reads	15	362	758	2043.5	98,050	531.25	1136.88	2.14
Replies	0	4	8	16	3405	15.34	31.98	2.08
Followers	0	5	14	49	10,127	54.52	456.76	8.38
Threads	1	1	2	5	595	5.84	20.12	3.44
Breast cancer forum								
Reads	14	297	504	854	90,783	368.81	719.18	1.95
Replies	0	8	15	26	1017	21.47	23.57	1.08
Followers	0	15	46	166	2627	219.26	474.06	2.16
Threads	1	1	3	12	5118	43.34	317.59	4.53
Diabetes consultation forum								
Reads	38	812	1203	1813	95,905	1065.60	1342.66	1.26
Replies	0	4	8	14	796	11.44	14.99	1.31
Followers	0	0	0	0	466	1.10	7.63	6.93
Threads	0	2	4	14	3862	20.60	94.14	4.57

[a]CV: coefficient of variation; CV=SD/mean.

Next, we plotted the frequency distribution of the number of reads per thread in each forum, as shown in Figure 1. We also adopted the K-S test (H_L=0, H_B=0, and H_D=0), from whose results we can determine that the number of reads per thread follows the gamma distribution [49]. When we plotted the frequency distribution of the number of replies per thread, we found that they exhibited an obvious long-tail phenomenon, which suggests that the number of replies per thread follows the power-law distribution. We further found that the log-log distribution of the number of replies per thread was noticeably curved, as shown in Figure 2, according to which we determined that the number of replies per thread follows an exponential distribution. The values of all fitted exponential curves were above 0.94 (=0.946, =0.958, and =0.971). Through similar analysis procedures, we further found that both the number of threads a user is involved with (=0.978, >=0.964, and =0.970) and the number of followers of a user (>=0.989, =0.962, and =0.990) followed the power-law distribution, and the values of all fitted power-law curves were above 0.96.

Figure 1. The distribution of the number of reads per thread in each of the forums (lung cancer, breast cancer, and diabetes consultation). For better visualization, the horizontal axis only shows the number of reads per thread up till 5000, since such threads hardly exist.

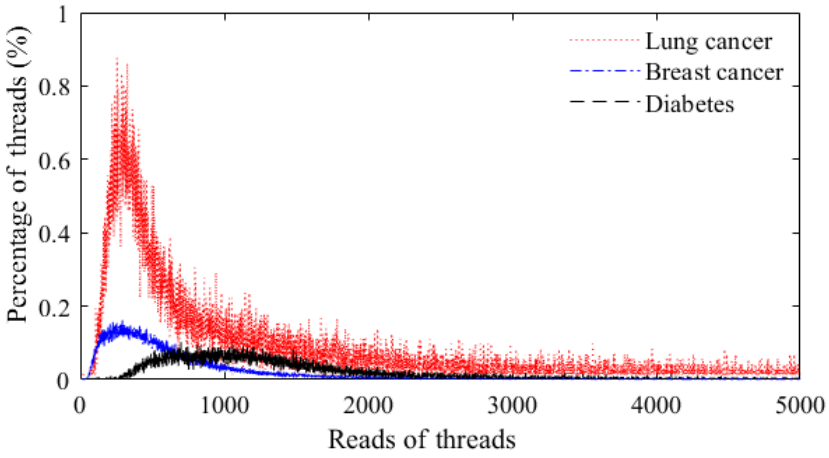
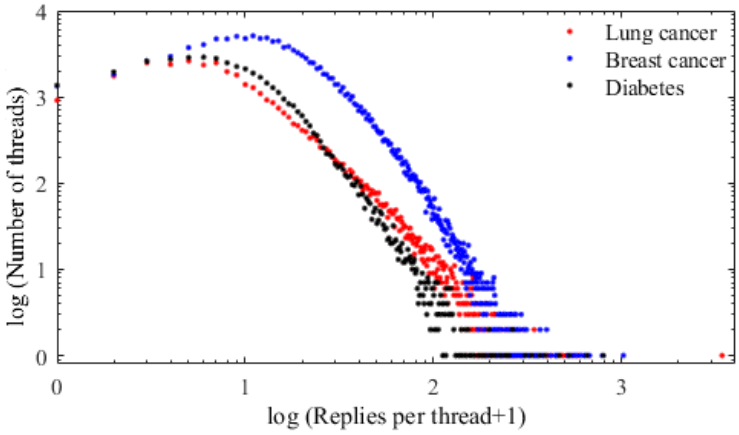


Figure 2. The log-log distribution of the number of replies per thread in each of the forums (lung cancer, breast cancer, and diabetes consultation).



User Activities

We explored key user activities on each forum. To understand user stickiness in a community, we analyzed online user activities. Community managers can adopt different strategies and incentives to improve their user experiences based on the behaviors of these users. Figure 3 shows the percentage of all posts (threads plus replies) created on each day of the week. These forums had similar trends, especially between the LCF and DCF. Besides, most online question and answer communities or vertical knowledge sharing communities are more active on weekdays than weekends, presumably due to users' conscious work-life balance choices. By counting the frequency of posting for each day of the week by month and drawing a box plot, it can also be seen that users in the LCF and DCF post more frequently during the week and are more active during the week than on the weekend (Multimedia Appendix 1). The same conclusion was found in nononline health communities such as the DISboards [36,50] and Tianya community [15,51].

Figure 3. Percentage of all posts on each day of the week.

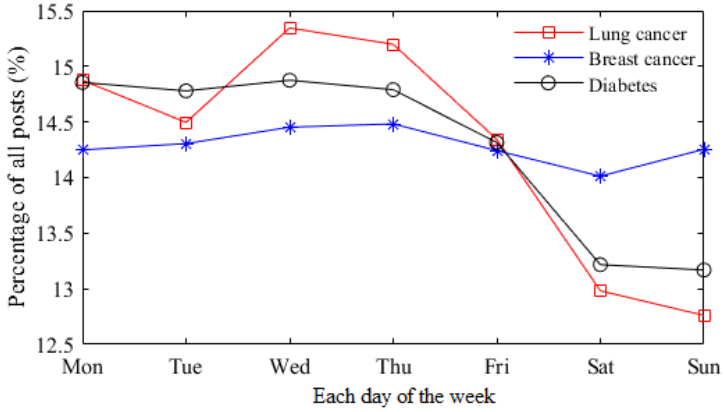


Figure 4 shows the percentage of threads and replies created at each hour of the day. In each forum, the number of posted threads and replies increased significantly from 4 AM. In terms of the posting time of each thread, the number of posted threads in the BCF had a peak around 8 AM, which declined in the middle of the day, followed by a second peak around 9 PM. Both the LCF and DCF showed 3 peak posting moments at around 10 AM, 4 PM, and 9 PM, with their least active posting moment at around 6 PM. Furthermore, there was high similarity (ρ=0.927; P<.001) between the LCF and DCF in terms of the relative thread posting frequencies during each hour of the day. The numbers of posted threads among the 3 disease forums around 12 PM and 6 PM were less than those at other moments in the day, presumably due to the common lunch and dinner hours observed for the 2 moments of the day. Most users became active from 7 PM after dinner, and activity peaked again at 9 PM, after which the number of posted threads gradually declined as it approached bedtime. We also found that the number of posted threads in the LCF and DCF peaked around 2 or 3 hours before the Chinese mealtime (12 PM and 6 PM), likely because diabetic patients pay more attention to their diet to control blood glucose. Similarly, lung cancer affects the digestive function of patients, which may cause decreased appetite. Especially in the advanced stage of lung cancer, it is indispensable to control and adjust the diet. Therefore, most users consulted about diet in advance, leading to a significantly increased number of posted threads. Due to the different dietary behaviors of breast cancer patients and those of patients with the other 2 diseases, the active posting periods of the BCF were different from those of the other 2 forums. In terms of the posting time of replies, we found that the relative frequencies of posting for threads and replies during each hour of the day were highly positively correlated in each forum. Table 3 shows the Spearman correlation test results. Similarly, the least numbers of replies were posted at around 12 PM and 6 PM, and this was also due to their overlap with common mealtimes.

Figure 4. Percentage of threads (A) and replies (B) at each hour of the day.

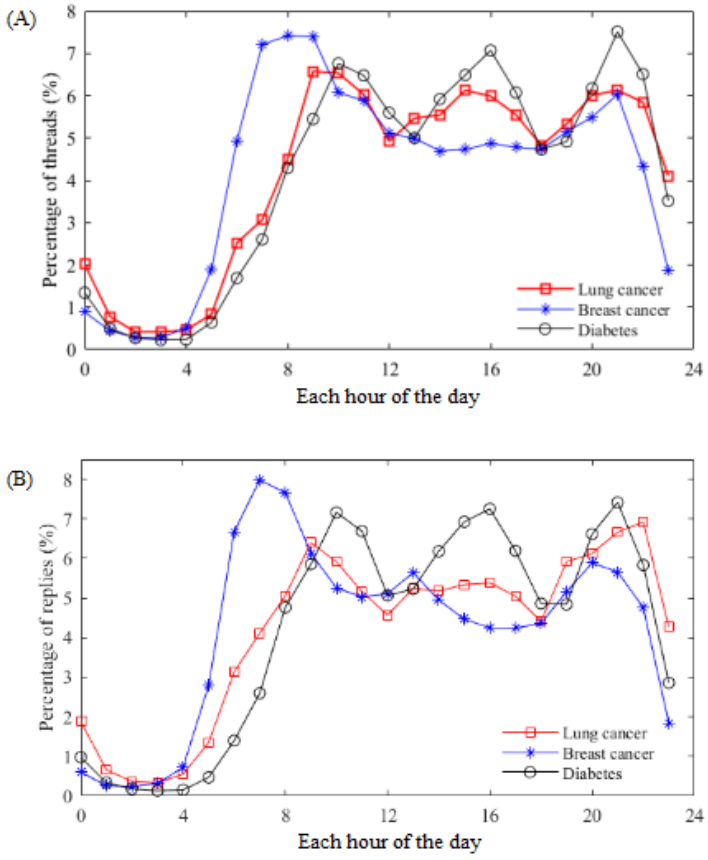


Table 3. Spearman rank correlation coefficients of the relative frequencies of posting for threads and replies during each hour of the day in each forum.

Forum	ρ	*P* value
Lung cancer forum	0.911	<.001[a]
Breast cancer forum	0.914	<.001[a]
Diabetes consultation forum	0.976	<.001[a]

[a]Significantly correlated using the significance level of .01 (2-tailed test).

Social Network Structure

The social network structure graph visually presents the node relationship matrix of the network [41]. Table 4 shows the characteristics of the aggregated social network based on users' reply postings on each forum. In the structure graph, each node represents a user in the community, and edges are directed links formed by replies between users. The average clustering coefficient of all 3 forums was lower than those of Facebook (0.519), Flickr (0.313), and LiveJournal (0.330) [52]. A higher global clustering coefficient indicates that there was a closer connection between users where friends tend to find each other through their mutual friends [36]. In a network, user degree is the sum of out-degree and in-degree of a user. In our data set, all users had at least one thread due to the crawling strategy used at the time of data acquisition. In the LCF, 39.7% (8968/22,610) of users had a user degree of 1; in other words, more than 39% of users only had a post. In the BCF, 51.4% (15,886/30,901) of users had a user degree of 1, and in the DCF, 24.7% (6604/26,751) of users had a user degree of 1. Most users had a relatively lower degree, while only a few users had higher degrees. Figure 5 shows the user degree distribution in each forum, from which we visually noticed that the degree distribution of each forum follows the power-law distribution. To quantitatively verify this finding, we treated the number of user degrees as an independent variable and the number of users as a dependent variable to fit a power-law curve. The resulting fitted curves were $y=3.571x^{-1.330}$ (R^2=0.9976) for the LCF, $y=3.253x^{-1.056}$ (R^2=0.9946) for the BCF, and $y=3.873x^{-1.445}$ (R^2=0.9683) for the DCF. The fitting degree of the power-law curves for each forum was nearly perfect, indicating that the distribution of user degrees on each forum follows the power-law distribution and $r>1$, which further shows that the underlying social network is a typical scale-free network. Many studies have reached the same conclusions that other OHCs [20,29,53], such as Facebook [52] and Weibo [54], are also typical scale-free networks. We additionally calculated the average clustering coefficient and the average path length of

the equivalent ER random networks for each forum. We randomly generated 50 sets of equivalent ER random networks, calculated σ according to equation (4), and calculated the mean value of 50 sets of σ as the final judgment coefficient. The final calculation results showed that all were far greater than 1 (=517.15, SD_L=13.31; =275.23, SD_B=13.02; and =525.18, SD_D=14.38), demonstrating that all 3 forums exhibited the small-world effect.

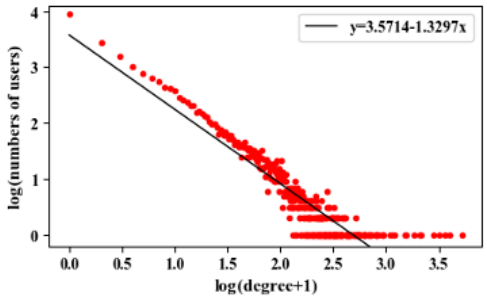
Table 4. Characteristics of each aggregated social network.

Characteristic	Lung cancer forum	Breast cancer forum	Diabetes consultation forum
Number of nodes	22,610	31,909	26,751
Number of edges	183,175	739,620	223,077
Average node degree	8.10	23.179	8.34
Network diameter	10	8	11
Average clustering coefficient	0.130	0.179	0.130
Average path length	3.494	3.011	3.967
Percentage of high-degree users[a]	3.1% (697/22,610)	9.1% (2906/31,909)	2.6% (697/26,751)
Percentage of low-degree users[b]	66.6% (15,050/22,610)	67.0% (21,382/31,909)	49.3% (13,057/26,751)

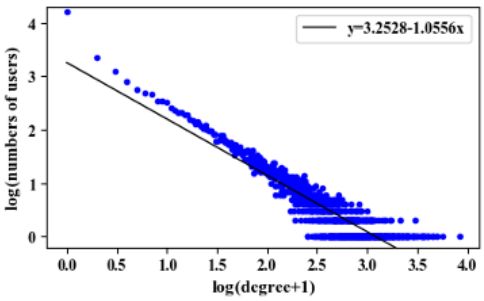
[a]Percentage of users with degrees higher than or equal to 100.

[b]Percentage of users with degrees lower than or equal to 5.

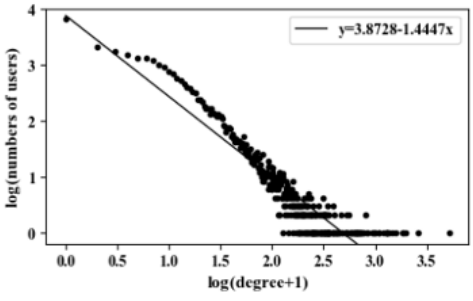
Figure 5. The total degree distribution of users for each of the aggregated social networks. (A) Lung cancer forum (LCF); (B) Breast cancer forum (BCF); (C) Diabetes consultation forum (DCF).



(A) LCF: y=3.5714-1.3297x



(B) BCF: y=3.2528-1.0556x



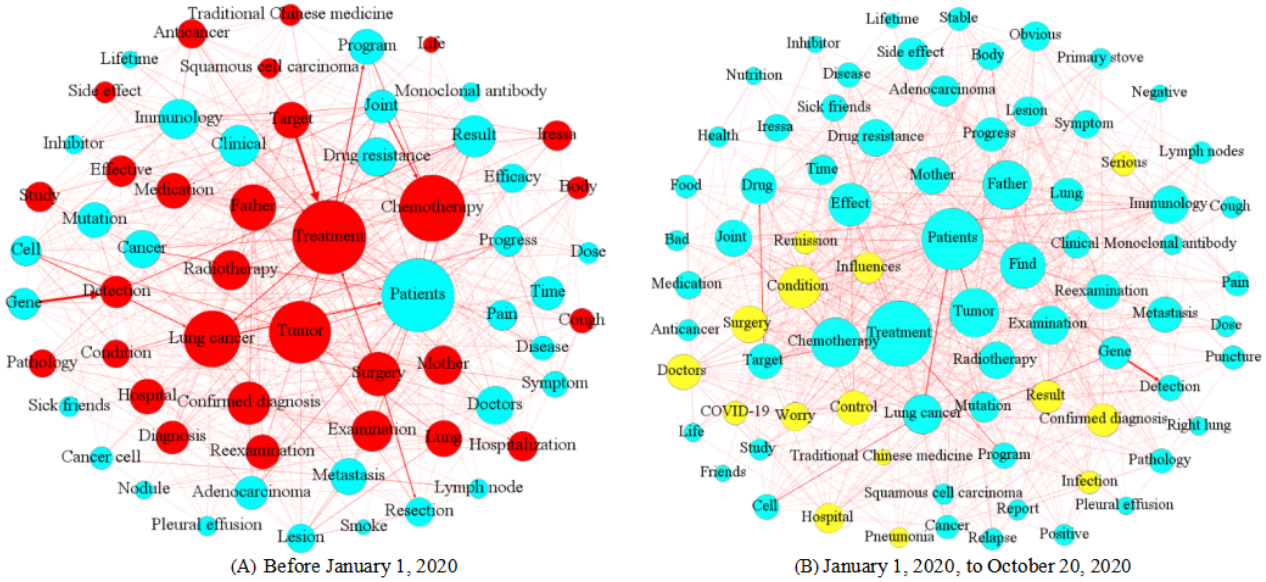
(C) DCF: y=3.8728-1.4447x

Next, we analyzed the dynamic evolution characteristics of the 3 social networks from their initial establishment days (LCF: November 15, 2013; BCF: August 25, 2015; DCF: September 1, 2005) to 2020 (Multimedia Appendix 2). The results indicated that the numbers of nodes and edges increased yearly since the creation of the LCF. With the development of the community, the number of users was gradually increasing accompanied by more active user behaviors. Since the establishment of the BCF in 2015, its number of nodes has been increasing yearly, indicating that new users constantly join the forum. In the meanwhile, the number of edges in the network increased in the beginning and reached a peak in 2017, and then declined afterwards. Such a trend line shows that the disease forum reached its most active period in 2017. Since the creation of the DCF back in 2005, the numbers of nodes and edges of the forum had been steadily increasing until its peak in 2015, after which the activity of the community continuously declined, in particular between 2018 and 2019. The "sleeping rate" or "loss rate" of users in the forum was noticeable. More concretely, the number of nodes active in 2019 was 21% (943/4510) of that in 2015, while the number of edges active in 2019 was only 5% (2731/55,741) of that in 2015, and both statistics indicate an apparent recession phase of the forum.

Analyzing UGC Using WKNs

Due to the COVID-19 outbreak in 2020 and its likely impact on user behaviors on OHCs, we divided our observation window into 2 periods (one before January 1, 2020, and another afterwards). By comparing user behaviors between these 2 periods, we analyzed whether the UGC changed notably due to the disease outbreak. We extracted the first 200 key phrases from the UGC of each of the 3 forums. In the preprocessing, we first filtered away keywords with no factual information according to peer literature, as well as merged synonym keywords [25,37]. For each forum, we subsequently constructed its corresponding WKN according to the construction method of the WKN model discussed above. Figures 6-8 show the resulting WKNs for each forum (Multimedia Appendix 3 for the detailed pictures). A larger node of a keyword in the WKN is associated with a greater weight of the keyword, implying more attention received by the keyword. Applying the criterion, we detected significant keywords in each forum, for example, the keywords "treatment" and "chemotherapy" in the LCF, according to Figure 6A. A darker color of the connection link between 2 keywords was associated with a higher co-occurrence frequency between these keywords. For example, the keyword "target" most frequently co-occurs with the keyword "treatment" in the LCF; thus, the connection link between these nodes has the darkest color in the forum's corresponding WKN as shown in Figure 6A. The dense connection of a keyword indicates that the keyword co-appears with many other keywords in a sentence, for example, the keywords "treatment" and "patients" in the LCF, according to Figure 6A.

Figure 6. Two separate weighted knowledge networks constructed for the lung cancer forum for the analysis phases (A) November 15, 2013, to January 1, 2020, and (B) January 1, 2020, to October 20, 2020.



There were 8 major categories of theme feature classification strategies, namely, "etiology and pathological knowledge," "diagnosis and examination," "treatment," "disease management," "complications," "social life," "disease prevention," and "education and research," based on the classification of OHC information in PubMed literature [55]. According to these 8 classification strategies, the classification categories of the keywords of the 3 disease forums were judged, and the topic distribution was macroclassified, so as to determine the hot topics discussed in the content more clearly and quickly. There were 400 keywords in the 200 key phrases. In the LCF, the topic "disease treatment" (145/400, 36.3%) included keywords such as "treatment" and "chemotherapy;" the topic "examination and diagnosis" (118/400, 29.5%) included keywords such as "examination," "confirmed diagnosis," and "detection;" and the topic "social life" (38/400, 9.5%) included keywords such as "sick friends" and "life." In the BCF, the topic "disease treatment" (128/400, 32.0%) included keywords such as "treatment" and "chemotherapy;" the topic "examination and diagnosis" (119/400, 29.8%) included keywords such as "examination" and "confirmed diagnosis;" and the topic "social life" (58/400, 14.5%) included keywords such as "sisters,"

"foods," and "sport." In the DCF, the topic "disease treatment" (155/400, 38.8%) included keywords such as "control," "treatment," and "injection;" the topic "examination and diagnosis" (115/400, 28.8%) included keywords such as "examination," "confirmed diagnosis," and "hyperglycemia;" and the topic "social life" (31/400, 7.8%) included keywords such as "foods," "sport," and "sick friends." It can be concluded that the 3 forums had "disease treatment," "examination and diagnosis," and "social life" themes. However, we noticed that topics related to disease prevention were rarely discussed in all 3 forums. Table 5 shows the top 10 keywords appearing in each forum, which primarily focused on disease treatment and diagnosis. Given the fact that target users of Mijian are patients after diagnosis, users of the forum tend to discuss topics on disease treatment and examination more frequently. More specifically, in both forums on Mijian (LCF and BCF) users generally paid more attention to topics on disease reexamination, metastasis, recurrence, anticancer drugs, and drug side effects. In both the BCF and DCF, users paid more attention to topics on healthy diet, exercise, and disease management.

Table 5. Top 10 keywords in each of the 3 forums.

Period	Lung cancer forum top keywords	Breast cancer forum top keywords	Diabetes consultation forum top keywords
Before January 1, 2020	Treatment, patients, chemotherapy, tumor, lung cancer, father, mother, confirmed diagnosis, surgery, and examination	Breast cancer, patients, treatment, tumor, chemotherapy, cancer, surgery, influences, examination, and metastasis	Blood glucose, control, insulin, fasting, treatment, normal, examination, diabetes, patients, and detection
January 1, 2020, to October 20, 2020	Treatment, patients, chemotherapy, tumor, find, father, effect, mother, lung cancer, and condition	Breast cancer, patients, treatment, tumor, chemotherapy, find, influences, increase, surgery, and examination	Blood glucose, control, insulin, fasting, treatment, normal, examination, diabetes, patients, and detection

According to Figure 6A, hot topics on the LCF before January 1, 2020, mainly concentrated on lung cancer treatment, examination and diagnosis, and social life. The most salient topic on the LCF was disease treatment, because this topic had the greatest weight. This category mainly focused on topics including keywords such as "lung cancer treatment," "chemotherapy," "surgical treatment," "drug treatment," and "treatment effect." The most relevant topics for treatment were examination and diagnosis, including keywords such as "lung pain" and "cough symptoms," indicating that users discussed examination and diagnostic contents, as well as the treatment of lung cancer. Ego networks consist of a focal node known as the ego, and the nodes to whom the ego is directly connected to, called alters, with edges showing links between the ego and altars or between altars. Each alter in an ego network can have its own ego network, and all ego networks combine to form the social network. The red nodes in Figure 6A formulate 2 ego networks for the keywords "father" and "mother." We found co-occurrence relationships among the keywords "mother," "father," "confirmed diagnosis," "reexamination," and "surgery," and they appeared in the same thread, implying that many users were probably children who consulted and communicated online health information for their parents. In the BCF, the hot topics before January 1, 2020, mainly focused on factors such as breast cancer treatment, social life, examination and diagnosis, and disease management, as shown in Figure 7A, and users mainly focused on breast cancer treatment and chemotherapy, such as endocrine therapy and treatment effects. The green nodes in Figure 7A formulate an ego network for the keyword "children." We found the keywords related to "result," and "health" was associated with "children." Due to the particularity of breast cancer, patients considered some special factors such as the health status of the next generation. Different node colors were used to more clearly distinguish the ego network of a particular keyword. Since the ego networks of other nodes do not have obvious characteristic results, only the ego networks of nodes with characteristic results are discussed in the paper. In the DCF (Figure 8), the UGC mainly concentrated on topics like diabetes control and management, disease treatment, examination and diagnosis, and social life, including keywords such as "blood glucose control" and "diet control." Meantime, the examination and diagnosis of diabetes were mostly related to disease management. In Figures 6B and 7B, yellow nodes formulate the ego network for the keyword "COVID-19," relating to keywords such as "infection," "remission," "confirmed diagnosis," and "influences." Cancer patients represent one of the susceptible populations of COVID-19, who are more vulnerable to COVID-19 complications [56] and prone to experience severe events on exposure to COVID-19, such as admission to an intensive care unit or death. Thus, they should pay more attention to self-protection and social distancing. Since 2020, both in the LCF and BCF, users tended to mention COVID-19–related matters while discussing their medical issues. Meanwhile, the relative mention frequencies of other key topics did not change noticeably. Given that there were only a few threads (only 130 threads) posted on the DCF after 2020, no clear thematic change was observed to draw any qualitative conclusions regarding the response of its participants to COVID-19. Thus, no distinction was made.

Figure 7. Two separate weighted knowledge networks constructed for the breast cancer forum for the analysis phases (A) August 25, 2015, to January 1, 2020, and (B) January 1, 2020, to October 20, 2020.

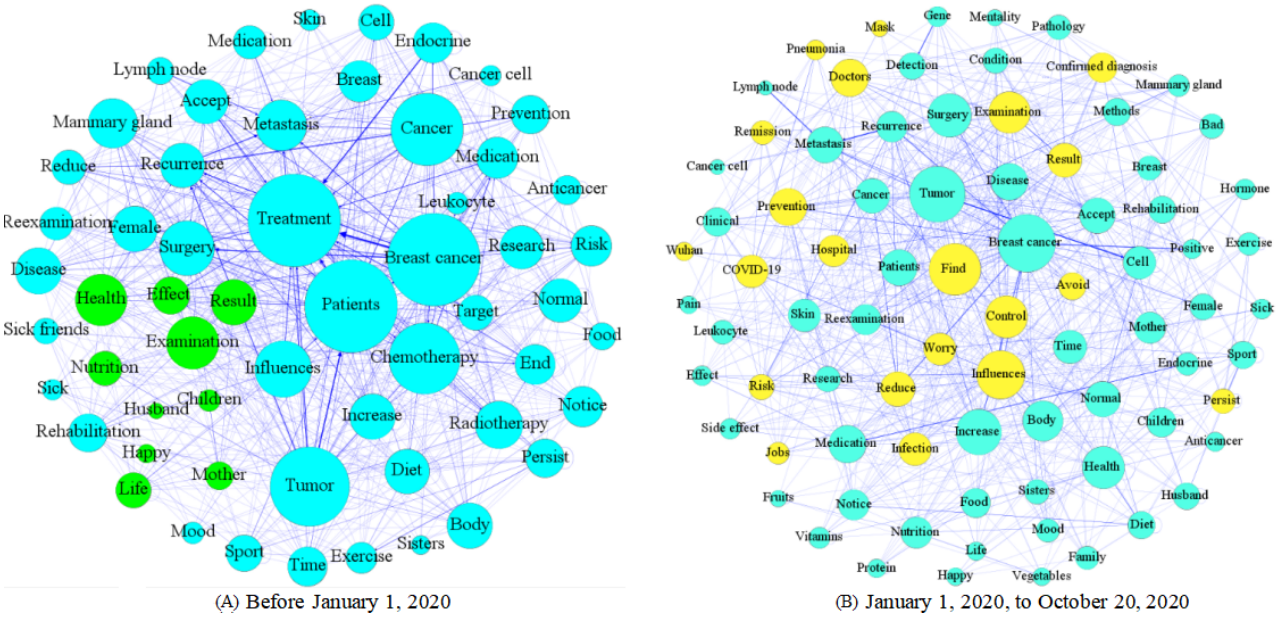
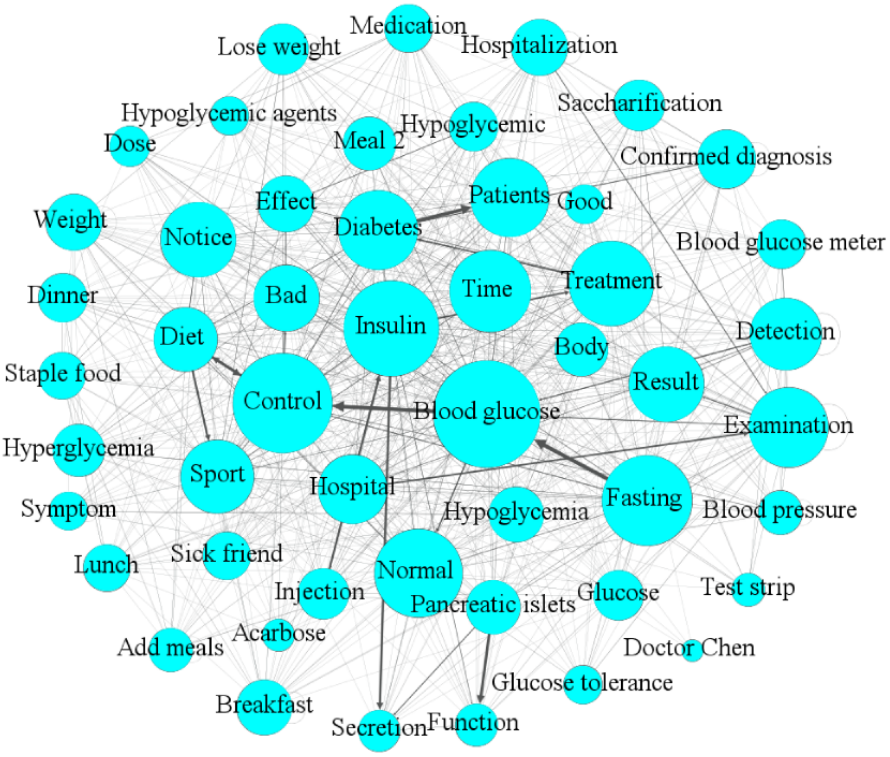


Figure 8. The weighted knowledge network constructed for the diabetes consultation forum during its full duration (September 1, 2005, to October 20, 2020).



Discussion

Principal Findings

This study carried out an in-depth analysis of the UGC and related online user behaviors of 3 large-scale OHCs in China. We utilized a variety of social network analysis methods and constructed a knowledge-sharing network for each OHC to study the evolution laws of the corresponding online community, discover characteristics of user behaviors, and uncover salient topics and their relations shared in the virtual community.

Since the existing research conducted on OHCs in China only examined a small-scale single disease forum, as shown in Table 1, that is, the number of data sets was less than 2000 threads and less than 10,000 replies [28,30,37,38], the scale was significantly smaller than that of analyses performed in Western countries [20,24,26], which severely undermines the reliability and comprehensiveness of the analysis findings. To meet the demand and fill the gap, we conducted thorough and extensive research on 3 representative disease forums selected from the 2 most popular Chinese OHCs. Over 80,000 users, 190,000 threads, and more than 2.8 million replies were crawled to reveal

the common traits and unique characteristics of user behaviors and UGC in these forums, which can better support our findings and represent the overall characteristics of OHCs in China. The results are discussed in detail below.

First, we found that the data of these 3 disease forums were polarized, and the underlying data distributions were certainly nonuniform. In these disease forums, the number of reads per thread followed gamma distribution (H_L=0, H_B=0, and H_D=0), and the number of replies per thread followed exponential distribution (=0.946, =0.958, and =0.971). However, the number of threads a user is involved with (=0.978, =0.964, and =0.970) and the number of followers of a user (=0.989, =0.962, and =0.990) both followed power-law distribution.

Second, users were more active during the weekdays than on weekends. The thread posting frequencies and reply frequencies in the abovementioned 3 forums had highly positive correlations between each other during each hour of the day. In particular, the LCF and DCF exhibited high temporal similarity (ρ=0.927; P<.001) in terms of the thread posting frequencies during each hour of the day. The numbers of threads and replies increased significantly from 4 AM, and the number of posted threads was relatively small in each forum around 12 PM and 6 PM. Because both lung cancer patients and diabetes patients need to pay attention to their diets, the number of posted threads in the LCF and DCF had a crest around 2 or 3 hours before the Chinese mealtime (12 PM and 6 PM).

Besides, the study showed that all 3 forums had the small-world effect (=517.15, SD_L=13.31, =275.23, SD_B=13.02, and =525.18, SD_D=14.38) and scale-free characteristics, and the user degrees followed the power-law distribution (=0.997, =0.994, and =0.968), while their global clustering coefficients were lower than those of international peer OHCs. According to the dynamic trends of the community networks, it was demonstrated that the LCF was still in the developing stage, the BCF needed to stimulate the activity of "zombie users," and the DCF needed to attract more new users and improve the retention rate of users.

Finally, we found that several hot topics were commonly shared among the abovementioned 3 disease forums, such as disease treatment, disease examination, diagnosis, and social life. The most relevant topics for treatment were examination and diagnosis, and many children consulted related information for their parents in the LCF. In the BCF, users paid more attention to the next generation's health, while in the DCF, users paid more attention to the detection of blood glucose and diet control. Furthermore, we noticed that in both the LCF and BCF, users tended to mention COVID-19–related matters while discussing their medical issues after the outbreak of the disease in 2020.

Limitations

There are few limitations in this paper. On one hand, although 2 influential OHCs in China (Mijian and Sweet Home) were selected for analysis in this study, the analysis results cannot be extended to all Chinese OHCs. On the other hand, this paper only focused on the characteristics of the overall social network structure, which did not distinguish the strong and weak connections between users and user roles. At the same time, this study only analyzed the topic content of 1 user, which did not consider replies or the topic type of a thread. Therefore, subsequent research should try to add weights to the connection edges between users to study the influence of users in social networks, or study the theme changes in different periods.

Conclusions

Our findings shed light on the basic characteristics of social networks, user behaviors, and UGC in Chinese OHCs. UGC in OHCs and related online user behaviors can be leveraged as an important source of information to gain insights on individual and population health conditions, which can be beneficial for users to understand hot topics in different forums and gain knowledge of health management. Despite the fact that OHCs are developing in China, it is indispensable to take measures to improve the retention rate and activity of users, increase user stickiness, analyze user behavior, and mine forum content themes. It is important to better mine potential content to provide users with useful information and knowledge. In conclusion, our research not only contributes to the understanding of the different characteristics of OHCs, but also helps to discover the salient topics and latent relations among these topics in each forum. Hence, effective, timely, and consistent mining and utilization of content can provide more valuable evidence for health providers and policymakers.

Acknowledgments

This research was supported by the Science and Technology Program of the Chinese Ministry of Science and Technology (grant number 2020AAA0106302), the National Natural Science Foundation of China (grant numbers 61876150 and 12026609), and the Science and Technology Program of the City of Xi'an, China (grant number XA2020-RKXYJ-0105).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Box plot of the frequency of postings for each day of the week by month.
[DOCX File , 55 KB - jmir_v23i12e19183_app1.docx]

Multimedia Appendix 2
The dynamic evolution characteristics of social networks in different years.
[DOCX File , 20 KB - jmir_v23i12e19183_app2.docx]

Multimedia Appendix 3
The detailed pictures of Figures 6-8.
[DOCX File , 2155 KB - jmir_v23i12e19183_app3.docx]

References

1. Internet Plus: Premier Li's new tech tool. The State Council of the People's Republic of China. URL: http://english.www.gov.cn/premier/news/2015/03/13/content_281475070887811.htm [accessed 2021-01-10]
2. Maloney-Krichmar D, Preece J. A multilevel analysis of sociability, usability, and community dynamics in an online health community. ACM Trans. Comput.-Hum. Interact 2005 Jun;12(2):201-232. [doi: 10.1145/1067860.1067864]
3. Open Medical and Healthcare Alliance. URL: https://www.omaha.org.cn [accessed 2021-01-10]
4. Action Plan for Further Improvement of Medical Services (2018-2020). National Health Commission of the People's Republic of China. URL: http://www.nhc.gov.cn/yzygj/s3594q/201801/9df87fced4da47b0a9f8e1ce9fbc7520.shtml [accessed 2021-01-10]
5. Patientslikeme. URL: https://www.patientslikeme.com/ [accessed 2021-01-10]
6. MyHealthTeams. URL: https://www.myhealthteams.com/ [accessed 2021-01-10]
7. Breastcancer.org. URL: https://www.breastcancer.org/ [accessed 2021-01-10]
8. BecomeAnEX. URL: https://www.becomeanex.org/ [accessed 2021-01-10]
9. Manyoubang. URL: http://www.manyoubang.com/ [accessed 2021-01-10]
10. Yi Xiang Network. URL: http://www.yx129.com/ [accessed 2021-01-10]
11. Mijian360. URL: http://www.mijian360.com/ [accessed 2021-01-10]
12. Sweet Home Diabetes Forum. URL: http://bbs.tnbz.com/ [accessed 2021-01-10]
13. Lymphoma Home. URL: https://www.house086.com/ [accessed 2021-01-10]
14. Tianya Community. URL: https://bbs.tianya.cn/ [accessed 2021-01-10]
15. Tieba.baidu. URL: https://tieba.baidu.com/ [accessed 2021-01-10]
16. Zhihu. URL: https://www.zhihu.com/ [accessed 2021-01-10]
17. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16(1):442-414 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
18. Dongxiang Z. Summary of the current situation of research on the online health community in China. Intelligence Work 2018;62(09):134-142. [doi: 10.13266/j.issn.0252-3116.2018.09.016]
19. Wu J, Liu G, Hu X. An Overview of Online Medical and Health Research: Hot Topics, Theme Evolution and Research Content. Data Analysis and Knowledge Discovery 2019;3(4):2-12. [doi: 10.11925/infotech.2096-3467.2018.1063]
20. Wang X, Zhao K, Street N. Analyzing and Predicting User Participations in Online Health Communities: A Social Support Perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
21. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. Computers in Human Behavior 2017 Oct;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
22. Fernandes LDS, Calado C, Araujo CAS. Social networks and health practices: influence of a diabetes online community on adherence to treatment. Cien Saude Colet 2018 Oct;23(10):3357-3368 [FREE Full text] [doi: 10.1590/1413-812320182310.14122018] [Medline: 30365855]
23. Li Y. Research on Influencing Factors of Medical Information Sharing Willingness of Patients in Online Medical Community. Wanfangdata. 2015. URL: https://d.wanfangdata.com.cn/thesis/ChJUaGVzaXNOZXdTMjAyMTA1MTkSB0Q3NTQxNDYaCGRqZG50NjNo [accessed 2021-10-19]
24. Wang X, Zhao K, Cha S, Amato MS, Cohn AM, Pearson JL, et al. Mining User-Generated Content in an Online Smoking Cessation Community to Identify Smoking Status: A Machine Learning Approach. Decis Support Syst 2019 Jan;116:26-34 [FREE Full text] [doi: 10.1016/j.dss.2018.10.005] [Medline: 31885411]
25. Jin B, Xu X. Research on Theme Features in Online Health Community. Intelligence Work 2015;59(12):100-105. [doi: 10.13266/j.issn.0252-3116.2015.12.015]
26. Della Rosa S, Sen F. Health Topics on Facebook Groups: Content Analysis of Posts in Multiple Sclerosis Communities. Interact J Med Res 2019 Feb 11;8(1):e10146 [FREE Full text] [doi: 10.2196/10146] [Medline: 30741640]
27. Bi Q, Shen L, Evans R, Zhang Z, Wang S, Dai W, et al. Determining the Topic Evolution and Sentiment Polarity for Albinism in a Chinese Online Health Community: Machine Learning and Social Network Analysis. JMIR Med Inform 2020 May 29;8(5):e17813 [FREE Full text] [doi: 10.2196/17813] [Medline: 32469320]
28. Wu J, Zhou L. The Study of Knowledge Sharing Network and Users Knowledge Interaction in Online Health Community. Information Science 2017;35(3):144-151. [doi: 10.13833/j.cnki.is.2017.03.026]

29. Zhao K, Wang X, Cha S, Cohn AM, Papandonatos GD, Amato MS, et al. A Multirelational Social Network Analysis of an Online Health Community for Smoking Cessation. J Med Internet Res 2016 Aug 25;18(8):e233 [FREE Full text] [doi: 10.2196/jmir.5985] [Medline: 27562640]
30. Shi J, Zhang B, Zhou L. Research on the Dynamic Evolution of Knowledge Sharing Network of Users in Health Question and Answering Community. Information Science 2019 Feb;37(02):77-82. [doi: 10.13833/j.issn.1007-7634.2019.02.013]
31. Lu Y, Wu Y, Liu J, Li J, Zhang P. Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community. J Med Internet Res 2017 Apr 07;19(4):e109 [FREE Full text] [doi: 10.2196/jmir.7087] [Medline: 28389418]
32. Zhu Y, Guan M, Donovan E. Elaborating Cancer Opinion Leaders' Communication Behaviors Within Online Health Communities: Network and Content Analyses. Social Media + Society 2020 Apr 02;6(2):205630512090947. [doi: 10.1177/2056305120909473]
33. Willis E, Royne MB. Online Health Communities and Chronic Disease Self-Management. Health Commun 2017 Mar 24;32(3):269-278. [doi: 10.1080/10410236.2016.1138278] [Medline: 27218836]
34. 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 12;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
35. Ogurtsova K, da Rocha Fernandes J, Huang Y, Linnenkamp U, Guariguata L, Cho N, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
36. Li Z, Chandler H, Shen H. Analysis of Knowledge Sharing Activities on a Social Network Incorporated Discussion Forum: A Case Study of DISboards. IEEE Trans. Big Data 2018 Dec 1;4(4):432-446. [doi: 10.1109/tbdata.2017.2749307]
37. Wu J, Wang Y, Li M, Cai S. Knowledge Discovery of Online Health Communities with Weighted Knowledge Network. Data Analysis and Knowledge Discovery 2019;3(2):108-117. [doi: 10.11925/infotech.2096-3467.2018.0619]
38. Wu J, Shi L. Study of the User Interaction Behavior in Online Health Community Based on Social Network Analysis. Information Science 2017;35(07):120-125. [doi: 10.13833/j.cnki.is.2017.07.021]
39. 39 Health Network. URL: http://www.39.net [accessed 2021-01-10]
40. Facebook. URL: https://www.facebook.com [accessed 2021-01-10]
41. Social network. Wikipedia. URL: https://en.wikipedia.org/wiki/Social_network [accessed 2021-01-10]
42. Milgram S. The Small-World Problem. Psychology Today 1967;1(1):61-67 [FREE Full text]
43. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998 Jun;393(6684):440-442. [doi: 10.1038/30918]
44. Humphries MD, Gurney K. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. PLoS One 2008 Apr 30;3(4):e0002051 [FREE Full text] [doi: 10.1371/journal.pone.0002051] [Medline: 18446219]
45. Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999 Oct 15;286(5439):509-512. [doi: 10.1126/science.286.5439.509] [Medline: 10521342]
46. Loeckx D, Slagmolen P, Maes F, Vandermeulen D, Suetens P. Nonrigid image registration using conditional mutual information. IEEE Trans Med Imaging 2010 Jan;29(1):19-29. [doi: 10.1109/TMI.2009.2021843] [Medline: 19447700]
47. Shannon CE. Prediction and entropy of printed English. The Bell System Technical Journal 1951;30(1):50-64. [doi: 10.1002/j.1538-7305.1951.tb01366.x]
48. Xi Y, Dang Y. A Method to Represent and Measure Personal Knowledge Stocks Based on Weighted Knowledge Network. Chinese Journal of Management 2007 Mar;4(1):28-31. [doi: 10.3969/j.issn.1672-884X.2007.01.005]
49. Karadağ Ö, Aktaş S. Goodness of fit tests for generalized gamma distribution. AIP Conference Proceedings 2016 Jun;1738(1):480119. [doi: 10.1063/1.4952355]
50. DISboards. URL: https://www.disboards.com [accessed 2021-01-10]
51. Yongliang Z, Xijin T. Online Behavior Analysis Based On Tianya Forum. Journal of Systems Science and Mathematical Sciences 2015;35(2):129-141. [doi: 10.12341/jssms12525]
52. Mislove A, Marcon M, Gummadi PK, Druschel P, Bhattacharjee B. Measurement and analysis of online social networks. In: IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. 2007 Presented at: 7th ACM SIGCOMM Conference on Internet Measurement; October 24-26, 2007; San Diego, CA, USA p. 29-42. [doi: 10.1145/1298306.1298311]
53. Durant KT, McCray AT, Safran C. Social network analysis of an online melanoma discussion group. Summit Transl Bioinform 2010 Mar 01;2010:6-10 [FREE Full text] [Medline: 21347134]
54. Luo Y, Zhou Z, Li L, Zhang H, Qin D. Social Network Disciplines Analysis Based on Power-law Distribution. Computer Engineering 2015;41(7):299-304. [doi: 10.3969/j.issn.1000-3428.2015.07.057]
55. Zhang J, Zhao Y. A user term visualization analysis based on a social question and answer log. Information Processing & Management 2013 Sep;49(5):1019-1048. [doi: 10.1016/j.ipm.2013.04.003]
56. Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, China Medical Treatment Expert Group for COVID-19. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J 2020 May;55(5):2000547 [FREE Full text] [doi: 10.1183/13993003.00547-2020] [Medline: 32217650]

Abbreviations

BCF: breast cancer forum
DCF: diabetes consultation forum
ER: Erdös–Rényi
LCF: lung cancer forum
NLP: natural language processing
OHC: online health community
PMI: point-wise mutual information
UGC: user-generated content
WKN: weighted knowledge network

Edited by R Kukafka; submitted 11.01.21; peer-reviewed by X Qin, X Luo; comments to author 10.02.21; revised version received 24.02.21; accepted 12.08.21; published 15.12.21.

Please cite as:
Lei Y, Xu S, Zhou L
User Behaviors and User-Generated Content in Chinese Online Health Communities: Comparative Study
J Med Internet Res 2021;23(12):e19183
URL: https://www.jmir.org/2021/12/e19183
doi:10.2196/19183
PMID:34914615

Original Paper

Impact of Medical Blog Reading and Information Presentation on Readers' Preventative Health Intentions: Mixed Methods, Multistudy Investigation

Kimberly A Taylor[1], BS, MA, PhD; William F Humphrey Jr[1], BBA, MBA, PhD

Department of Marketing & Logistics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
William F Humphrey Jr, BBA, MBA, PhD
Department of Marketing & Logistics
College of Business
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3053480148
Email: whumphre@fiu.edu

Abstract

Background: Medical blogs have become valuable information sources for patients and caregivers. Most research has focused on patients' creation of blogs as therapy. But we know less about how these blogs affect their readers and what format of information influences readers to take preventative health actions.

Objective: This study aimed to identify how reading patient medical blogs influences readers' perceived health risk and their intentions to engage in preventative health actions. Further, we aimed to examine the format of the medical blog and the reader's response.

Methods: We surveyed 99 university participants and a general-population, online panel of 167 participants. Both studies randomly assigned participants to conditions and measured blog evaluation, intentions for preventative health action, and evaluation of health risk and beliefs, and allowed open-ended comments. The second study used a different sample and added a control condition. A third study used a convenience sample of blog readers to evaluate the link between reading medical blogs and taking preventative health action.

Results: Across 3 studies, participants indicated a desire to take future preventative health action after reading patient blogs. Studies 1 and 2 used experimental scenario-based designs, while Study 3 employed a qualitative design with real blog readers. The 2 experimental studies showed that the type of blog impacted intentions to engage in future preventative health actions (Study 1: $F_{2,96}$=6.08, P=.003; Study 2: $F_{3,166}$=2.59, P=.06), with a statistical blog being most effective in both studies and a personal narrative blog showing similar effectiveness in Study 2, contrary to some prior research. The readers' perceptions of their own health risk did not impact the relationship between the blog type and health intentions. In contrast, in one study, participants' judgments about the barriers they might face to accessing care improved the fit of the model ($F_{2,95}$=13.57, P<.001). In Study 3's sample of medical blog readers, 53% (24/45) reported taking preventative health action after reading a health blog, including performing a self-check, asking a doctor about their health risk, or requesting a screening test. Additionally, these readers expressed that they read the blogs to follow the author (patient) and to learn general health information. All studies demonstrated the blogs were somewhat sad and emotional but also informative and well-written. They noted that the blogs made them appreciate life more and motivated them to consider taking some action regarding their health.

Conclusions: Reading patient blogs influences intentions to take future health actions. However, blog formats show different efficacy, and the readers' disease risk perceptions do not. Physicians, medical practitioners, and health organizations may find it useful to curate or promote selected medical blogs to influence patient behavior.

(J Med Internet Res 2021;23(12):e23210) doi:10.2196/23210

KEYWORDS

health blogs; patient blogs; preventative care; cancer; caregivers; perceived risk

Introduction

Background

Patients and caregivers rely on online blogs, social media posts, or online health communities to share information about their illness, treatment paths, or feelings about their condition [1]. Some share with a small circle of friends and family, while others make their blog posts public and shareable with a worldwide audience. Many studies have shown the therapeutic power that writing has for patients [2,3]. Some cancer organizations now even offer writing classes and activities for survivors [4], and many health care providers are writing blogs as well [5,6]. Comparatively less research has focused on the blog readers, however.

Over the years, the web has become more interactive and "participatory" with many peer-to-peer communication forms, such as social media, blogs, and wikis, making up much of what we do online [7]. Softwarefindr estimated there were over 505 million blogs online [8] in 2018. As of 2020, internet users create new posts every 0.5 seconds, while 77% of them say they read blogs regularly [9]. Though lifestyle topics are among the most popular categories [10], personal health and illness-related blogs have proliferated as many people go to the internet as their first source of health-related information [11].

As the Pew Internet Project explains, "peer-to-peer health care acknowledges that patients and caregivers know things—about themselves, about each other, about treatments—and they want to share what they know to help other people." We have a natural tendency to want to both seek and share information about our health, and technological advances have made this easier than ever [12]. The internet allows community interactions between individuals, including patients and caregivers, and it can be a resource for individuals to discover health information. Further, it has facilitated access to health information and emotional support [13,14]. Social media conversations also have proven effective in improving patient knowledge of their conditions and reducing anxiety [14], but research has demonstrated that individuals are more likely to consume social content instead of creating it [15]. In fact, while the internet has become more participatory, a relatively smaller group of "power users" contribute more than the average user [15].

Researchers studying cancer patient bloggers found that writing their own and reading other patients' stories affected these patients' perceptions of their illnesses and prognoses [16]. These cancer patients kept blogs to be remembered after their death, to release negative emotions and frustrations, and to share information about their experience with others, especially other patients. Additional research found that patients explore other patients' experiences, find community and belonging, and gain a sense of hope by reading and following survivor stories on social media [17-19]. A study of 5 women experiencing depression reported similar results, finding that the blogs helped the writers stave off feelings of seclusion, find community, and form bonds with readers [20]. Moreover, the author coined the term "narrative sandbox" to refer to these blogs as a "protected virtual space that allows bloggers to temporarily and experimentally add or remove different sections from their narrative" [20]. As such, these narratives are dynamic and changeable, with the blog readers becoming "active participants in the writing and rewriting of bloggers' depression narratives" [20].

Caregivers—usually parents of ill children, adult children caring for aging parents, or the patient's spouse—write many health blogs as well, and studies of cancer caregiving report the key motivations: "to report, explain, express, reflect, archive and advocate" [21]. Thus, these parent blogs provide information to readers by reporting events and explanations of medical terms while also expressing feelings, emotions, and reflections about events. A study of blogs by caregivers of dementia patients found similar themes, including social support and engagement, gathering and sharing information, reminiscing and building legacies, and altruism [22]. Just as the parents of children with cancer often have to advocate for their children, the family members of patients with dementia often feel drawn to activism and want to advocate for others. Some researchers argue that blogs about the end of life are understudied. In her analysis of 3 late-stage cancer blogs, Andersson [23] posits that the blogs provide useful language around illness, death, and dying, which in turn produces emotional responses in the reader. However, the paper also recognized the deficiencies in our language and feelings of powerlessness. The paper noted that the blogs often feature narratives of struggle and fighting the disease, positive thinking, and even magical thinking, but little direct discussion of death. This lack of adequate language for discussing death creates a meaningful bond between the writer and the reader in their "shared ineffability" or an inability to properly express their thoughts and feelings about death.

Blog Readers and Health Beliefs

We have seen that blogs about a treatment plan can enhance patient-provider relationships [24], and participation in online groups can help empower patients and counteract isolation [25]. Yet, researchers know comparatively less about the influence of these blogs on their readers. Much of the literature on reading medical blogs refers to medical students as the readers and how these blogs enhance medical education [26,27]. Although that is indeed a worthwhile goal, medical students are not the target of the present research.

One survey of cancer blog users identified 3 different clusters [28] that varied by their motivations for reading blogs, and one segment had the most behavioral change (seeking changes in care) due to the blogs. However, 59.29% of their sample were cancer patients, 31.86% were family or friends of cancer patients, and the remaining 6.19% were medical professionals, and this study did not explicitly differentiate between the characteristics of the blog content or between blog users who were the writers and those who were readers.

A study that explicitly studied blog readers found 4 reasons for reading blogs and linked these to 3 behavioral outcomes [29]. The 4 motivations for reading blogs included "affective exchange, information search, entertainment, and getting on the bandwagon," while outcomes included changing readers' opinions, reader-writer interaction, and spread of word-of-mouth communications to others. Thus, given that reading blogs can change one's opinions and motivate some actions, it is

reasonable to conclude that readers might seek to monitor their health or take preventative action upon reading an illness blog.

Some characteristics of an illness blog may produce greater behavioral change than others. According to the Stanford Encyclopedia of Philosophy, Aristotle presented 3 means to persuasion: the character or credibility of the speaker (*ethos*), the emotional response of the listener (*pathos*), or the argument itself (*logos*) [30]. In addition, Cialdini [31] proposed that people employ decisional heuristics, or shortcuts, to deal with the volume of information they encounter. One such heuristic is similarity, in which people trust and believe those they deem to be like themselves [32]. Past research also has associated social proof, another persuasion tool in which we follow what we see others doing, with persuasion and conforming to established norms [33]. Similarly, other persuasion tools of commitment and consistency by Resnik and Cialdini [33] have also been associated with compliance in behavior.

This research is interested in patient blogs rather than those written by medical professionals. Thus, we did not investigate the writer's credibility (ethos) due to specialized knowledge or expertise. Instead, the study considers the blog's emotional response (pathos) through a personal narrative, which may demonstrate similarity or social proof to the reader, and objective data in the form of cancer statistics (logos), which may be an indication of disease risk.

Past research found that personal stories (narratives) increased both the perceived risk of infection with the hepatitis B virus and the intention to vaccinate more than statistical evidence [34]. This research suggested that personal narratives are less prone to counterarguing or discounting by the reader, as messages inconsistent with one's prior beliefs have been shown to be [35]. Thus, the researchers proposed: "narratives are hence expected to be superior in conveying personal health risks than statistical evidence" [35]. Similar work found statistical data were less persuasive in an alcohol education message [36], while other research found anticipated emotions more strongly predicted influenza vaccination than perceived risk [37]. The presentation of objective information caused people to rely on a peripheral cue—the expertise of the source—"so that an expert communicator induced greater persuasion than did a source with lower expertise" [38]. Readers would not be likely to consider a patient to be an expert source, and thus, perhaps, patient blogs might be less persuasive when written with statistical data than when written as a personal story.

Objective

While researchers have explored patient use of writing, we know less about the readers of these writings, and these studies aimed to fill this gap. The research's primary objective was to determine whether medical blog readers intend to take some health action specifically after reading a health blog (objective 1).

Specifically, we posited that blog type will significantly affect intentions to take some personal health action, with a *personal narrative* eliciting higher intentions than a blog with *statistical data*. Thus, the secondary objective was to determine specific characteristics of the most likely blogs to produce these health changes (objective 2). The final objective was to determine whether the blog reader's perceived health risk will mediate the effect of blog type on behavioral intentions (objective 3). That is, do readers believe themselves to be at higher risk after reading a health blog and then take some preventative action? To explore the 3 objectives, we employed a combination of quantitative and qualitative methods across 3 studies.

Methods

Study Design

All research was conducted in the United States using Qualtrics XM survey software to randomize treatment conditions for Studies 1 and 2. Study 3 was a qualitative survey with no treatment conditions.

Studies 1 and 2

The first study was a single-factor design with 3 levels of medical blog types. This study presented participants with 1 of 3 cancer blog excerpts—a personal narrative (story), a blog with statistical data about disease prevalence, and a more general cancer blog. The second study replicated those 3 conditions with a different sample population and added a control condition (with no blog excerpt).

Study 3

A third study of real blog readers provided some external validity and assessed actual preventative health actions taken (rather than hypothetical, intended actions) after reading a patient blog. This study utilized a qualitative survey with no intervention or experimental conditions.

Participants

Study 1

The participants were 99 students in an academic research subject pool at a large southeastern public university. The student participants received a few extra credit points in a college class in exchange for participating in the study. Potential participants saw a brief description of the study posted using SONA research panel software, where they could decide which studies they wished to complete. The completion rate was 97.1% (99/102) of the sample, while the remaining 2.9% (3/102) abandoned the online survey without answering any questions and thus were dropped from the final sample. The sample had an average age of 24.5 years and was 53% (53/99) female.

Study 2

Study 2 used a sample from the Amazon Mechanical Turk (MTurk) panel who were over 18 years old. Amazon workers have their choice of which studies they wish to complete and when. Participants included 167 Amazon master MTurk workers (past participants who have provided quality responses to other researchers) who received a US $2 payment for participation. An additional 22 participants began the survey and read the online consent form but answered no other questions; they were thus dropped from the final dataset, resulting in an 88.4% (167/189) participation rate. Respondents were 56.4% (94/167) female with an average age of 44.8 years.

Study 3

Participants were a convenience sample of actual medical blog readers recruited through posts to the authors' social media, which were then reposted or forwarded in a snowball sample. Although 59 participants began the survey, 8 participants said they had not read any medical blogs, and 3 participants were not sure or did not respond to the question. An additional 3 participants responded that they had read a medical blog but did not answer any further questions. Thus, the final sample contained responses from 45 participants. Table 1 summarizes the sampling, recruitment, and analysis methods employed in the 3 studies.

Table 1. Summary of participants in the 3 studies that focused on blog intention investigations.

Characteristics	Study 1 (n=99)	Study 2 (n=167)	Study 3 (n=45)
Type of sample	University research panel	Amazon MTurk master respondent panel	Convenience and snowball sample of medical blog readers
Average age (years), mean (range)	24.5 (18-46)	44.8 (21-76)	Not collected
Location	Southwestern US research university	United States	United States
Research tools	Qualtrics XM Survey Software	Qualtrics XM Survey Software	Qualtrics XM Survey Software
Development of tools	Adaptation of existing scales and creation of blog posts	Adaptation of existing scales and creation of blog posts	Blog evaluation scales repeated from studies 1 & 2 and additional questions
Prior relationship with researcher	Primarily no (4 of the 99 were enrolled in the researcher's course)	No	Some—convenience snowball sample
Blinded	No—identifying information collected in a separate file from participant response data, per institution IRB[a] guidelines	Yes	Yes
Statistical analysis			
Quantitative data	One-way ANOVA[b] (SPSS)	One-way ANOVA, stepwise regression (SPSS)	Simple means and % reported
Qualitative data	Hand coding of open-ended responses	Hand coding of open-ended responses	No coding—qualitative responses reviewed and summarized
Female, n (%)	53 (53.3)	94 (56.4)	Not collected
Education	Current undergraduate students	Not provided by panel	Not collected
Recruitment	University research panel software: Sona Systems	Posted on Amazon MTurk[c] open studies to eligible panelists	Recruitment via social media and snowball sample
Percentage of original sample that completed the study, %	97.1[d]	88.4[e]	76.3[f]

[a]IRB: institutional review board.

[b]ANOVA: analysis of variance.

[c]MTurk: Mechanical Turk.

[d]N=102.

[e]N=189.

[f]N=59.

Intervention and Instruments

Study 1

Survey instruments were developed primarily by adapting existing materials and scales. The researchers adapted the blog posts used in Studies 1 and 2 from a publicly available, widely shared blog post about someone with skin cancer. This public blog excerpt was adapted and revised to create the 3 conditions (cancer statistics added to create that condition, for example). All blog posts were similar in total word length and readability. The Qualtrics survey software randomly assigned participants to 1 of 3 experimental conditions (for Study 1), which differed only in which blog post was presented to participants. The 3 blog types were a personal narrative (focused exclusively on a patient's personal story), a general cancer story, and a third statistics condition (that incorporated data and statistics about melanoma, rather than a personal story). All sample blog posts featured the same photo of a young woman and were approximately the same length.

After first reading the informed consent form and consenting to participate, subjects saw the sample blog post and answered an open-ended question asking their thoughts about the blog. This open-ended question was followed by an evaluation of the blog itself in terms of readability, informativeness, and interest on 7-point bipolar scales (all specific scale items and instruments are included in Multimedia Appendix 1).

Next, participants indicated how likely they were to engage in several preventative health actions (see a doctor for a skin check, monitor your skin yourself for any changes, use sunscreen daily, use sunscreen when going to the beach, or ask a doctor about cancer risk) on 7-point scales (from 1 = not at all likely to 7 = extremely likely). Then, respondents completed health beliefs model scales [39]; scales were reworded only to relate to skin cancer or skin checks rather than vaccination. These scales comprised measures of perceived barriers to seeking care, perceived benefits of skin checks, perceived susceptibility to skin cancer, and perceived severity of the effects of skin cancer. Although not intended to be predictive, we included these scales to explore potential relationships between constructs.

Finally, participants completed locus of control (loc) scales measuring internal, external-other (in which powerful others control events), and external-chance (in which events are due to fate or luck) loci of control [39]. The survey concluded with a few personal health questions (whether they used sunscreen and how often, whether they have ever had a skin check with a dermatologist, and whether they had a prior diagnosis of skin cancer). On average, participants spent 652.85 seconds (nearly 11 minutes) on the survey.

Study 2

Study 2 was conducted to replicate the significant effects found in Study 1 with a larger, general adult (nonstudent) sample population. Moreover, we added a control condition in which participants responded to the health intention, health beliefs, and locus of control measures but were not shown any blog post. Thus, for this study, the Qualtrics software randomly assigned participants either to 1 of the 3 blog-type conditions or to the no blog condition.

All study manipulations and measures were identical to those of Study 1, except for the addition of the control condition. In the control condition, instead of presenting a blog sample, we asked participants where they usually get their health information and their evaluations of that source. This design kept the survey to a similar length and provided participants a task to complete before answering the health intention measures. On average, respondents spent 618.17 seconds (10.3 minutes) on the survey.

Study 3

The first 2 studies were hypothetical, scenario-based designs. Thus, we also conducted a concise survey of real blog readers to explore our research objectives and assess the effect of medical blogs on actual preventative behaviors (rather than intentions). After asking respondents to think about a medical blog that they had read or followed, the survey asked who wrote the blog they read, whether they took any personal health-related actions after reading the blog, and if so, what actions. Next, the survey asked their reasons for reading the blog, what they liked or did not like about reading the blog, and for an evaluation of the blog writing itself. The researchers developed all measures for this study. These respondents completed the 7-question survey in an average of 212.9 seconds (3.55 minutes).

Statistical Analysis

We analyzed the quantitative scale data in Studies 1 and 2 with one-way analyses of variance (ANOVAs) and a simple regression model using SPSS version 26. Study 3 did not include an intervention; thus, the only quantitative data reported for that study were descriptive means and percentages.

The qualitative data from Studies 1 and 2 were analyzed through a coding process. First, researchers developed a list of codes for the open-ended, written evaluations of the blog posts based upon prior literature and an initial review of a sample of responses. All thought listings were downloaded to a spreadsheet and separated by respondent. Next, 2 coders, blind to both experimental condition and research objectives, independently coded all responses manually in the spreadsheet. Across the data from Studies 1 and 2, there was nearly 70% agreement (69.3%) between the 2 coders with a κ value of 0.63 [40]. This was deemed a sufficiently strong level of agreement, and a third coder resolved discrepancies.

Ethics

All studies were approved by the Social and Behavioral Sciences Institutional Review Board (IRB) at a large, public university in the southeastern United States (IRB 20-0088). Participants for all studies completed an online consent form prior to beginning the study and were free to omit any questions or leave the study at any time.

Study 1 was not blinded, as participant information was required for awarding the compensation (of the extra class credit). However, in accordance with the university's IRB guidelines, the identifying information was saved to a separate file and not connected with the experimental responses. Studies 2 and 3 were blinded. Amazon MTurk (used for Study 2) provides only a user ID, and participant identities are not revealed to researchers. The final qualitative survey (Study 3) used Qualtrics anonymous response settings and did not capture any identifying information.

Results

Study 1

The blog excerpts were rated overall as being easy to read (mean 5.84, SD 1.61), easy to understand (mean 6.29, SD 1.07), well-written (mean 6.02, SD 1.34), informative (mean 5.64, SD 1.31), emotional (mean 6.16, SD 1.07), and interesting (mean 5.82, SD 1.37). Importantly, these evaluations did not differ between the 3 blog conditions (all $F_{2,98}$ values <1.3, P>.30 for one-way ANOVAs). That is, the different blog types were judged to be equally informative, easy to read, and emotional. Thus, any significant effects of blog type on the health intentions could not be attributed to these characteristics. Table 2 summarizes the individual means for the evaluation of the blog posts, as well as the behavioral intentions.

Table 2. Study 1 blog evaluations and behavioral intentions.

Characteristics	Blog type			Overall (N=99), mean (SD)
	Personal (n=34), mean (SD)	General (n=38), mean (SD)	Statistics (n=27), mean (SD)	
Blog evaluation				
Easy to read	5.76 (1.58)	5.92 (1.60)	5.81 (1.71)	5.84 (1.61)
Easy to understand	6.32 (1.01)	6.18 (1.21)	6.41 (0.97)	6.29 (1.07)
Well-written	6.06 (1.41)	5.97 (1.33)	6.04 (1.32)	6.02 (1.34)
Informative	5.50 (1.33)	5.68 (1.25)	5.74 (1.16)	5.64 (1.31)
Emotional	6.26 (1.05)	6.26 (1.01)	5.89 (1.16)	6.16 (1.07)
Interesting	5.59 (1.44)	6.00 (1.21)	5.85 (1.49)	5.82 (1.37)
Behavioral intention				
Skin check by a doctor	4.26 (1.83)	4.55 (1.75)	5.04 (1.85)	4.59 (1.81)
Skin self-check	5.29 (1.43)	5.71 (1.43)	6.30 (0.869)	5.73 (1.35)
Sunscreen daily	4.15 (2.05)	4.47 (2.19)	5.30 (2.15)	4.59 (2.16)
Sunscreen at beach	5.85 (1.87)	6.18 (1.45)	6.70 (0.669)	6.21 (1.49)
Asked a doctor about skin cancer risk	4.58 (1.82)	4.84 (1.94)	6.11 (0.974)	5.10 (1.79)

Next, we examined whether blog type affected the readers' health intentions. First, the set of health intention measures proved to be reliable as one scale (Cronbach α=.79), with an overall mean of 26.2 (SD 6.4). The effect of blog type on health intentions surrounding skin cancer prevention was significant ($F_{2,96}$=6.08, *P*=.003). However, the direction was not as predicted, as both patient blog types (personal: mean 23.97, SD 1.26; general: mean 25.76, SD 6.10) elicited lower intentions to take preventative health action than did the blog that presented statistics without a personal story (mean 29.44, SD 4.54; *P*=.001, *P*=.02, respectively). Table 2 summarizes the means of the individual health intention measures for all conditions.

On the individual health intention measures, we observed significant differences between blog conditions only for self-monitoring of one's skin ($F_{2,98}$=4.47, *P*=.01) and asking a doctor about one's cancer risk ($F_{2,97}$=6.9, *P*=.002). On both measures, the blog with statistics led to greater preventative health intentions than either of the other 2 blog conditions (personal narrative and general cancer–related.) The 2 measures about using sunscreen demonstrated *P* values of .11 and .08, respectively, while the first measure about seeing a doctor for a skin check did not differ between blog conditions (*P*=.26).

The third research proposition was that blog type would influence health beliefs. First, we assessed the reliability of all health beliefs subscales, including the perceived barriers to accessing health care (Cronbach α=.63), perceived susceptibility (Cronbach α=.68), and perceived severity subscales (Cronbach α=.69; the perceived benefit of getting a skin check was a single-item measure, not a multi-item scale). However, the reliability of the perceived barriers scale was higher (Cronbach α=.65) without the second item ("a skin check could have unpleasant side effects"), and the reliability of the perceived severity subscale was higher without the first item ("skin cancer may lead to serious health problems"; α=.712 without that item). Thus, the final reported scales did not include those items. Moreover, the external-others locus of control scale was used in full (α=.78), while both the internal loc and external-chance loc scales had greater reliability if some items were dropped (α_s=.78 for resulting scales; internal loc removed the first 2 items, and chance loc removed the third item); thus, the reduced scale means are reported in the following sections. Table 3 summarizes the health belief subscales and locus of control scales (described in subsequent sections), and all scale items are listed in Multimedia Appendix 1.

Table 3. Study 1 locus of control.

Locus of control (LOC)	Blog type			Overall, mean (SD)	Number of items
	Personal, mean (SD)	General, mean (SD)	Statistics, mean (SD)		
Perceived barriers	14.06 (5.18)	12.46 (3.97)	11.08 (4.12)	12.62 (4.57)	4
Perceived benefits	5.19 (1.38)	5.32 (1.44)	5.85 (1.32)	5.42 (1.398)	1
Perceived susceptibility	8.75 (3.99)	8.76 (3.23)	9.96 (4.38)	9.09 (3.83)	3
Perceived severity	17.24 (5.84)	18.53 (5.49)	20.00 (6.08)	18.50 (5.81)	5
Internal LOC	18.70 (4.77)	18.84 (4.68)	18.81 (4.21)	18.79 (4.54)	4
Powerful others LOC	24.39 (7.21)	24.79 (6.64)	25.52 (5.08)	24.86 (6.41)	6
Chance LOC	13.91 (6.23)	15.03 (5.88)	14.52 (4.53)	14.51 (5.63)	5

Neither the perceived susceptibility to skin cancer ($F_{2,96}$ =.965 P=.38) nor the perceived severity subscales ($F_{2,97}$=1.7, P=.19) varied by blog condition. The perceived benefits of skin checks also did not differ by condition. However, perceived barriers to screening did significantly differ by blog condition ($F_{2,96}$=3.373, P=.04), with the personal narrative showing greater barriers than the statistics blog. We did not predict a priori any effects of blog condition on internal, powerful others or chance locus of control, and none were found (all F values <1, all P>.70).

Given that the perceived barriers to accessing care differed by blog type, we incorporated that scale alone into a regression model. A simple linear regression analysis regressing both blog type and perceived barriers to care onto the health intentions scale yielded an overall model that was significant ($F_{2,95}$=13.57, P<.001), as well as both variables that were significant (β_{blog}=.227, t_{92}=2.38, P=.02; $\beta_{perbarrier}$=–.36, t_{92}=–3.74, P<.001). Thus, adding the perceived barriers to care does not eliminate the effect of blog type on health intentions. Moreover, retaining the blog type in the regression model improved the model R^2 from .179 to .226 (Table 4).

Table 4. Regression analysis results.

Model factor	Estimate	SE	95% CI LL[a]	95% CI UL[b]	P value
(Constant)	10.951	2.679	5.630	16.271	<.001
PerBarScale	0.510	0.136	0.240	0.781	<.001
Blog version	–1.859	0.780	–3.409	–0.310	.02

[a]LL: lower limit.

[b]UL: upper limit.

Next, we turned our attention to the analysis of the open-ended question about the blog. The codes used for this qualitative analysis are shown in Table 5, along with sample responses and the number of responses coded into each category. The most-reported codes for the Study 1 sample were sadness/sympathy/empathy (code 1), makes one appreciate life more (code 7), informative (code 4), well-written (code 5), and makes one think of someone with cancer (code 8).

Table 5. Coding of open-ended responses and sample respondent statements in Study 1.

Open-ended question code category	Sample statements	Number of responses
1. Sadness/sympathy/empathy	I feel saddened by the writer's situation and an urge to help by any means necessary.	17
2. Feel guilty (eg, to be alive and well when others are not)	I feel guilty that I do not appreciate my life enough.	1
3. Feel worried/concerned for oneself	I feel nervous when I think about this sort of thing happening.	0
4. Informative	I think that it is informative about the feelings of people with cancer, which is something that you do not hear about often.	11
5. Well-written	I think that Natalie is an excellent writer who is expressing her true feelings about having cancer.	9
6. Motivates behavior (eg, will make an appointment, do a skin check)	After reading this, I really felt like I want to go and do my annuals, which is pending.	2
7. Makes you appreciate life more	That life is a gift. And we need to leave it with meaning and love what we do. Be thankful.	13
8. Thought of someone with cancer	I'm thinking of all of the people that I know who have had cancer. I have lost so many family members to cancer.	9
9. Did not want to read/would not read	I immediately wanted to close the page when I saw the topic. I feel sick to my stomach. This is not a topic I would ever want to read about.	4
10. No new information/heard it all before	There are many blogs like that, and I have heard many similar ones.	2

The distribution of codes across the 3 blog versions was also significantly different ($X^2_{16,68}$=28.54, P=.03). Both personal blog types elicited great feelings of sadness and sympathy (code 1), with the personal narrative blog also being considered well-written (code 5) and the general blog making readers appreciate their own life more (code 7). Participants most often cited the statistics blog as being informative (code 4). Notably, 3% (2 out of the 68 who provided a response to the open-ended question) stated, without prompting, that the blog made them intend to take some personal medical action.

Study 2

In this sample, the blog excerpts were again rated overall as being easy to read (mean 6.28, SD 1.17), easy to understand (mean 6.65, SD 0.724), well-written (mean 6.10, SD 1.28), informative (mean 5.47, SD 1.43), emotional (mean 5.91, SD 1.26), and interesting (mean 5.64, SD 1.62). There were few differences in these evaluations across the 3 blog types, except for how understandable ($F_{2,123}$=3.45, P=.04) and emotional ($F_{2,123}$=3.40, P=.04) the blogs were. The blog with the very personal story was rated more understandable than the general blog (P=.01), and the personal story was rated more emotional than the statistics-focused version (P=.01). Table 6 summarizes the evaluations of the blog posts.

Table 6. Study 2 blog evaluations.

Characteristic	Blog type			Overall (N=167), mean (SD)
	Personal (n=42), mean (SD)	General (n=41), mean (SD)	Statistics (n=41), mean (SD)	
Easy to read	6.14 (1.42)	6.12 (1.21)	6.59 (0.706)	6.28 (1.17)
Easy to understand	6.81 (0.505)	6.41 (0.974)	6.71 (0.559)	6.65 (0.724)
Well-written	6.26 (1.27)	5.85 (1.30)	6.17 (1.26)	6.10 (1.28)
Informative	5.38 (1.58)	5.24 (1.46)	5.78 (1.22)	5.47 (1.43)
Emotional	6.19 (1.37)	6.02 (1.24)	5.51 (1.08)	5.91 (1.26)
Interesting	5.76 (1.57)	5.56 (1.67)	5.59 (1.63)	5.64 (1.62)

Next, we turned our attention to the primary research objectives—does reading a personal illness blog affect the reader's own health intentions (objective 1)? Specifically, is there an effect of the blog type on these health intentions (objective 2)? First, the set of health intention measures proved to be reliable as a scale (Cronbach α=.80), with an overall mean of 24.48 (SD 6.94). The experimental condition's effect on health intentions was not significant in the overall ANOVA ($F_{3,166}$=2.59, P=.06). However, looking at the planned comparisons revealed that all blog conditions led to higher health intentions than the control (no blog) condition, but there were no significant differences between the 3 experimental blog conditions. Specifically, the personal (mean 25.21, SD 6.64; P=.03) and statistical (mean 25.95, SD 6.81; P=.01) blog versions led to higher overall health intentions than did the no blog control (mean 22.02, SD 7.27) condition. The general blog

(mean 24.49, SD 6.84) condition did not significantly differ from either the control or the other blog conditions.

One-way ANOVAs conducted on the individual health intention measures showed a significant difference existed only for daily sunscreen use ($F_{3,166}$=4.39, P=.005), with the only other noteworthy comparison observed for getting a skin check by a doctor ($F_{3,166}$=2.06, P=.11). Post hoc tests showed that, for daily sunscreen usage, all blog conditions indicated greater intentions than the no blog condition (P=.002, P=.004, and P=.005, respectively), while both the personal story and statistics blog versions led to greater intentions to get one's skin checked by a doctor than the no blog condition (P=.04 and P=.03, respectively). Table 7 summarizes the health intention measures by experimental condition.

Table 7. Study 2 behavioral intentions.

Characteristic	Control (no blog), mean (SD)	Blog type			Overall (N=167), mean (SD)
		Personal (n=42), mean (SD)	General (n=41), mean (SD)	Statistics (n=41), mean (SD)	
Skin check by a doctor	3.42 (2.04)	4.33 (2.09)	3.88 (2.05)	4.39 (2.02)	4.00 (2.07)
Skin self-check	5.65 (1.36)	5.74 (1.74)	5.71 (1.33)	5.95 (1.52)	5.76 (1.49)
Sunscreen daily	3.28 (2.33)	4.69 (2.12)	4.61 (2.02)	4.56 (1.86)	4.28 (2.16)
Sunscreen at beach	5.95 (1.75)	6.33 (1.41)	5.93 (1.88)	6.54 (1.31)	6.19 (1.61)
Ask a doctor about skin cancer risk	3.81 (1.92)	4.12 (1.82)	4.37 (1.91)	4.51 (2.03)	4.20 (1.92)

As we did collect demographic data on age and gender in this study, we also evaluated whether either variable affected the behavioral intentions, and they did not. The average response by men (mean 23.91, SD 6.71) on the health intentions scale did not significantly differ from that of the female-identifying respondents (mean 25.11, SD 7.00; t_{158}=1.1; P=.27), nor were there any gender differences on any of the specific items (all P values >.13). A regression of age on behavioral intention also was not significant ($F_{2,161}$=1.3, P=.28).

Respondents in this study also completed the health beliefs and locus of control scales, which we used to consider our third research objective about readers' perceived risk. Although there is some criticism of how the Cronbach α statistic is used and disagreement as to what constitutes an acceptable reliability, values above .6 or .7 are frequently reported as acceptable [41].

All scales here demonstrated Cronbach α>.62: perceived barriers (Cronbach α=.79), perceived susceptibility (Cronbach α=.63), perceived severity (Cronbach α=.63), internal locus of control (Cronbach α=.77), external-other locus of control (Cronbach α=.82), and external-chance locus of control (Cronbach α=.84). Moreover, there were no differences on any of these scales by experimental condition (all F<1.2, P>.20). This implies that neither the reading of a blog (versus no blog) nor the different types of blogs affected any of the respondents' health beliefs or their loci of control. Post hoc comparisons did, however, reveal a significant difference in evaluations of external-chance locus of control between the statistics blog (mean 16, SD 6.24) and the no blog (control) condition (mean 19.12, SD 6.41; P=.04); no other differences emerged. Table 8 summarizes these results.

Table 8. Study 2 locus of control.

Locus of control (LOC)	Control	Blog type			Overall (N=167), mean (SD)	Number of items
		Personal (n=42), mean (SD)	General (n=41), mean (SD)	Statistics (n=41), mean (SD)		
Perceived barriers	17.1 (6.78)	15.9 (6.07)	16.32 (7.57)	14.68 (5.88)	16.01 (6.61)	5
Perceived benefits	5.52 (1.04)	5.74 (1.15)	5.44 (1.34)	5.88 (1.05)	5.64 (1.16)	1
Perceived susceptibility	9.62 (3.86)	10.10 (3.68)	9.83 (3.73)	10.17 (4.02)	9.93 (3.68)	3
Perceived severity	24.30 (5.64)	24.60 (5.07)	23.55 (5.55)	23.73 (4.98)	24.05 (5.28)	6
Internal LOC	28.88 (4.89)	28.64 (6.00)	29.60 (5.90)	28.93 (5.09)	29.00 (5.46)	6
Powerful others LOC	23.56 (7.11)	24.36 (6.23)	23.29 (7.14)	23.93 (6.63)	23.79 (6.73)	6
Chance LOC	19.12 (6.44)	18.31 (6.92)	17.66 (7.66)	16.00 (6.24)	17.78 (6.87)	6

The number of open-ended thoughts coded into each category for Study 2 are summarized in Table 9. Overall, the most common evaluations were feelings of sadness/sympathy, that the blog was informative, it made them think of someone they know, and that there was no new information presented. Comparing across conditions, we see that the most common responses again varied by blog type ($X^2_{16,121}$=35.49, P=.003). For the 2 personal blog versions, the top codes were for expressions of sadness/sympathy/empathy (code 1) and thinking of someone they knew with cancer (code 8). The blog with

statistics evoked far different responses, with informative (code 4), well-written (code 5), and motivates behavior (code 6) being the most reported. Across the 3 blog types, 5 responses explicitly expressed worry for oneself, and 8 expressed an intent to be proactive regarding their health (scheduling a doctor's appointment or getting a skin check, for example.) Thus, without prompting, nearly 11% (13/124, 10.5%) expressed concern about or desire to act regarding their health because of reading one sample blog entry.

Table 9. Coding of Study 2 open-ended responses.

Open-ended question code category	Number of responses
1. Sadness/sympathy/empathy	41
2. Feel guilty (eg, to be alive and well when others are not)	0
3. Feel worried/concerned for oneself	5
4. Informative	22
5. Well-written	15
6. Motivates behavior (eg, will make an appointment, do a skin check)	8
7. Makes you appreciate life more	1
8. Thought of someone with cancer	18
9. Did not want to read/would not read	3
10. No new information/heard it all before	18

Study 3

This nonexperimental study surveyed actual self-described readers of real medical blogs. Most of the blogs read by our study participants were written by patients (27/45, 60%), followed by the patient's spouse or caregiver (7/45, 16%), a medical professional (6/45, 13%), or the patient's parent (5/45, 11%). The participants read the blogs for various reasons, chiefly to stay up to date with their friend or family member's condition (23/45, 51%) and to provide support for their friend (21/45, 47%). Many also read the blogs to gain information about a health condition more generally (11/45, 24%) or to learn information that may be pertinent for their own health (11/45, 24%). Percentages sum to more than 100% because they were able to select more than one reason. They found the blogs easy to read (mean 6.34, SD 1.14), easy to understand (mean 6.25, SD 1.19), well-written (mean 6.13, SD 1.50), informative (mean 6.23, SD 1.38), and interesting (mean 6.24, SD 1.87). Blogs were also considered somewhat emotional (mean 4.97, SD 0.83; all measured on 1 to 7 Likert-type scales).

The central research question for this study (and the paper's first research objective) was whether readers took any personal health action or changed any behaviors based upon reading the blog, and, indeed, 24 respondents (24/45, 53%) reported doing so. These responses are summarized in Table 10. Of participants who reported taking action, 29% (7/24) scheduled a doctor's appointment, while others reported requesting a cancer or other health screening (7/24, 29%); performing a self-exam, such as a skin or a breast check (5/24, 21%); or asking a doctor about their own risk (4/24, 17%). Additionally, 54% (13/24) reported making other health changes, including taking supplements, making lifestyle modifications, purchasing cancer insurance, doing additional research, or making donations to disease research.

Table 10. Preventative health actions taken by participants in Study 3 (n=45).

Preventative health action taken	Responses, n (%)
Scheduled a doctor's appointment	7 (16)
Requested cancer or other health screening	7 (16)
Performed a self-exam	5 (11)
Asked a doctor about their own health risk	4 (9)
Other (made other health or lifestyle changes)	13 (29)

Participants cited being able to stay updated, especially without having to impose on their friend or family member for continued updates, as the greatest motivator for reading the medical blog. Additional motivations included learning specific ways they could support the person, feeling connected to the person, and hearing positive news when that was shared. Several also cited that the blog writer was very transparent, real, and open, which they appreciated, or used humor. Comments also corroborated prior findings about the use of medical statistics, with others citing how the writer explained detailed scientific information so that it was understandable for readers, sometimes prompting them to do additional research. Several participants with health challenges also said reading the blogs provided some relief to know that others were going through the same things and helped confirm their own treatment choices. Most said there was nothing they did not like or found difficult about reading the

blogs, but 24% (11/45) of participants described the experience as very emotional or sad to read, notably if the person's condition deteriorated or a treatment was unsuccessful. Those participants appreciated the blog and found it worthwhile to read, despite being sad or emotional at times. Interestingly, a few mentioned that the information was not always correct or that they would take it more as a suggestion for lifestyle modification or a reason to do more research, rather than as medical advice.

Discussion

Principal Findings

We conducted 2 scenario-based, experimental studies with different sample populations and 1 qualitative survey of real blog readers to explore the research objectives. All 3 studies confirmed the first research objective, that reading medical blogs was associated with intentions to take preventative health actions, though each study contributed unique findings. Studies 1 and 2 examined the specific characteristics of blogs that led to greater health intentions (objective 2), but neither perceived risk nor severity of skin cancer mediated these effects (objective 3). Study 3 confirmed the first objective with real, rather than intended, behaviors.

Study 1

This study confirmed that blog type significantly affects intentions to take preventative health action. We predicted that people would be motivated to take health actions upon reading a medical blog and that a highly personalized blog would produce the greatest health intentions. However, we discovered the reverse—the blog with melanoma statistics produced the most significant intention to take preventative health action, higher than either of the other blog conditions, which did not differ. Moreover, the statistics blog also yielded the lowest perceived barriers to accessing care compared with the other blog types. Including both blog type and perceived barriers to accessing care improved the predictive ability of the regression model on health intentions. Thus, blog type significantly affected perceived barriers, which impacted health intentions, but blog type's direct effect on behavioral intentions retained its significance. We conclude that medical blogs can produce intentions to engage in protective health behaviors, partially through an effect on perceived barriers to care.

Study 2

Study 2 corroborates and extends the findings of Study 1 by demonstrating that reading a medical blog excerpt led to greater intentions to take preventative health actions compared with a baseline (from a control condition in which participants did not read a blog entry), and this was true for all the blog versions we tested. Further, both a highly personal blog and a statistical blog led to greater personal health intentions than the control condition. In contrast, the more general cancer blog produced a moderate level of health intention that was not statistically different from either the control or the other blog conditions.

However, there were no differences by blog type on any of the health beliefs. Thus, perceived health risk—either susceptibility or severity—again did not mediate the effect of blog type on personal health intentions. Also, blog type and blog reading, in general, did not affect perceptions of either the barriers to or the benefits of skin checks. Interestingly, the statistics-oriented blog did reduce estimates of locus of control attributed to chance.

Study 3

The final study assessed whether reading medical blogs induced their readers to take any preventative health action in a real-world survey rather than a hypothetical scenario. For a majority of the respondents, it did. While the respondents primarily read the blogs in order to keep up to date with the patient's condition and provide support, they also found the writing emotional and informative, and many readers took personal health action as a result of reading the blogs. Reading medical blogs inspired people in many ways; they made medical appointments or requested health screenings, conducted additional research on their own, made donations to organizations, or made other lifestyle changes.

General Discussion

Researchers have long known medical blogs have demonstrable patient benefits, but less was known about the impact on their readers. The top intended health actions across our studies included skin checks and consulting with a physician about cancer risk, both of which are important preventative health actions that can help ensure that a skin cancer is caught early in a more treatable stage [42]. Furthermore, while our first 2 studies measured hypothetical health intentions, much past literature incorporating Ajzen's Theory of Planned Behavior [43] to health care has found intentions to be predictive of behaviors such as attending health screenings, engaging in healthy eating, or participating in regular physical exercise [44-46]. Thus, the increased intentions are likely to translate into actions, and Study 3 provides some support for that conjecture.

Some of our results also ran counter to predictions. The blog readers in our studies demonstrated a higher propensity to take preventative health actions when blog posts focused on statistics rather than personal stories. Whereas some prior literature showed personal stories to be most persuasive [34], our Study 1 found the statistics-oriented blog to be the most effective, and Study 2 found it to be as persuasive as the personal narrative. Because cancer can be scary and overwhelming to contemplate, there is a natural tendency to believe that cancer patients might have had either genetic or environmental reasons for their cancer diagnosis [47], to distance oneself from the situation and downplay one's own personal risk. Accordingly, it is possible that a personal story feels more specific to the writer and less relatable to the reader. In the past, people even avoided those with cancer out of fear they might "catch" the disease themselves [48]. This reduced relatability could minimize Cialdini's [49] similarity persuasion tool of the personal narrative. If one feels that the patient's story is unique and personal, then the reader may feel sadness and sympathy for the patient (as many of our participants in all 3 studies did) but may not feel motivated to take personal action because they read the story. And in fact, almost 47% of our Study 3

respondents said they did not engage in any preventative health actions after reading the blog.

In retrospect, perhaps the statistical blog increased the social proof aspect of persuasion by illuminating for our study participants that the condition is more widespread than they may have thought [31]. Providing some additional support for this conjecture, Study 1 results showed that the statistics blog reduced perceived barriers to screening compared with the personal story blog. Further, Study 2 showed that the statistics blog reduced the locus of control attributed to chance. Things that occur due to chance or fate, by definition, cannot be prevented by one's own actions. This effect could play a role in increasing one's intention to take personal preventative action if the perceived portion of cancer risk due to chance or fate is reduced. Perceived behavioral control—or one's belief in their ability to perform a behavior—is a central tenet in the Theory of Planned Behavior and has been found to be an important driver of intentions in health care [50]. The prospect of these potential mechanisms deserves further research to clarify our understanding.

We further explored whether the blog post conditions would differ significantly from a control (baseline) condition without any blog post. Indeed, all blog types we studied (general, statistics-focused, and a personal story) demonstrated greater health intentions as compared with the condition with no blog. Specifically, the personal story and statistics-focused blogs were associated with intention to use sunscreen, while all 3 blog types were associated with greater intent to schedule a medical visit for a skin check compared with the control condition. In our real-world follow-up study, we found that the desire to stay up to date on a patient's condition or find general information motivated survey participants to read medical blogs. As a result of reading these patient posts, more than 53% of respondents indicated that they took a preventative medical action (skin checks, doctor appointments) or other health-related behavior (additional research, insurance purchase, starting supplements). Remarkably, none of the blogs impacted perceived severity of or susceptibility to disease. Thus, perhaps even those readers who may believe they are at low risk might benefit from reading health blogs and be spurred to act.

Limitations

Limitations of this research relate to natural concessions made in sampling. The first study used a research panel at a major research university, while the second study used a panel from MTurk. Thus, neither sample may be representative of US adults more generally. Nonetheless, both samples included a diverse age range and, despite skewing a bit more female, at least 44% male-identifying respondents. For the sake of parsimony and because we did not have specific research objectives about age or gender, Study 3 did not collect this or any other demographic data. Thus, we cannot compare the results from that study to prior literature that examined these variables. Further, we did not collect educational data for the second or third studies.

Despite limitations inherent with any panel, both experimental studies provided convergent findings and extended our understanding of the efficacy of different blog presentations compared with one another and compared with a baseline without blog content. We designed the final study to understand blog readers' actual preventative health actions and to increase the robustness of qualitative insights.

The third study used a relatively small convenience sample with snowball sampling to provide insights into real actions resulting from reading patient health blogs. Although qualitative research often precedes quantitative efforts, the first 2 experimental studies' findings piqued the researchers' curiosity to explore the research objectives further and gain additional insights with a sample of readers of nonhypothetical medical blogs. Moreover, while we could have included many other variables, we kept the Study 3 survey purposefully short to reduce the burden on participants who received no compensation for participating. This study nonetheless provides an important step in helping to make a connection between blog reading and personal preventative health actions.

Next, this research inquiry looked at specific outcomes related to skin cancer. It is possible that the most influential blog style may differ when investigating other conditions, such as healthy behaviors postbariatric surgery or prevention of the spread of infectious viruses such as COVID-19. We anticipate that a relationship between reading health blogs and intended health actions would exist, but the presentation style and influence of perceived risk might be different. Additionally, these studies are among the first of which we are aware to investigate the impact of reading patient blogs by social media participants. Patients frequently turn to social media and online communities for medical information and advice [51,52]; future research is needed to explore additional predictors and outcomes in this area.

Conclusions

Blogs from patients undergoing health care treatments have become quite common. This article corroborates that, despite the emotional connection driven by personal story blogs, posts focusing on statistics related to the condition may be more effective at driving readers' preventative health intentions. This increased intention to perform preventative health action occurs regardless of the reader's perception of risk. Further, the current research clarifies that blog posts in general, regardless of format, are more effective at driving some preventative health intentions compared with merely thinking about where one gets health information. This finding indicates that reading patient-created content may be beneficial regardless of the reader's own risk for the health condition. Reading patient blogs may also impact perceived barriers to accessing care or perceptions of the part of our health risks that may be within our control (and not attributable to chance), leading to greater intentions to take preventative health actions. Finally, we gained insight that reading blogs, particularly those written by patients, were motivated primarily by a desire to support the patient, keep up with the patient, and learn more about the patient's condition. Some readers did indicate that they did further research and did not take blog information at face value, but a majority reported engaging in protective health action due to their blog reading.

Although much research exists on the benefits of patients' writing and journaling during health care treatment, the present research provides a foundation for future studies on patient and

health outcomes from reading health blogs. As people seek information online about medical conditions and preventative options, patient-generated content will appear between content created by and for medical professionals; health care practitioners cannot assume patients will only read edited content from medical professionals. Understanding what type of content presentation is most effective in encouraging positive health actions may guide health care providers, patient coordinators, and patient therapists to guide recommended content styles with the greatest impact. In the case of the present studies, the use of informative statistics was the most effective in driving these intentions. This research could also inspire future studies in other health specialties to understand how these results may generalize across medical conditions and treatments.

Acknowledgments

The authors acknowledge the readers of the first author's own "ProfessorMom" CarePages breast cancer blog and the PostHope blog of Shannon B Rinaldo, PhD (1973-2020), which documented her 5-year treatment at MD Anderson for glioblastoma IV, as inspiring and informing this research. They also thank Sphurti Sewak and Nuket Serin for their research assistance with coding the qualitative responses and reviewers AX, AY, and AN for their helpful comments on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study 1 and 2 measures.
[DOCX File , 19 KB - jmir_v23i12e23210_app1.docx]

References

1. Yu H, Wang Y, Wang J, Chiu Y, Qiu H, Gao M. Causal effect of honorary titles on physicians' service volumes in online health communities: retrospective study. J Med Internet Res 2020 Jul 09;22(7):e18527 [FREE Full text] [doi: 10.2196/18527] [Medline: 32673232]
2. Lepore SJ, Smyth JM, editors. The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-Being. Washington, DC: American Psychological Association; 2002.
3. Lepore SJ, Greenberg M, Bruno M, Smyth JM. Expressive writing and health: self-regulation of emotion-related experience, physiology, and behavior. In: Lepore SJ, Smyth JM, editors. The Writing Cure: How Expressive Writing Promotes Health and Emotional Well-Being. Washington, DC: American Psychological Association; 2004:99-117.
4. Lowry F. Cancer Patients Write to Heal in Unique Writing Workshop. Medscape. 2017 Aug 22. URL: https://www.medscape.com/viewarticle/884576 [accessed 2021-11-13]
5. Miller EA, Pole A, Bateman C. Variation in health blog features and elements by gender, occupation, and perspective. J Health Commun 2011 Aug;16(7):726-749. [doi: 10.1080/10810730.2011.551994] [Medline: 21432711]
6. Scheibling C, Gillett J, Brett G. Making the virtual rounds: the use of blogs by health-care professionals. Journal of Communication Inquiry 2017 Jul 24;42(1):48-66. [doi: 10.1177/0196859917722364]
7. Fox S, Madden M. Riding the Waves of "Web 2.0". Pew Research Center. 2006 Oct 5. URL: https://www.pewresearch.org/internet/2006/10/05/riding-the-waves-of-web-2-0/ [accessed 2021-11-13]
8. How Many Blogs are there in 2018? SaaS Scout. URL: https://www.softwarefindr.com/how-many-blogs-are-there/ [accessed 2021-11-13]
9. Blogging Statistics. 99Firms. URL: https://99firms.com/blog/blogging-statistics/#gref [accessed 2021-11-13]
10. Revealed: Which are the Most Popular Types of Blogs? WPBeginner. URL: https://www.wpbeginner.com/beginners-guide/which-are-the-most-popular-types-of-blogs/ [accessed 2021-11-13]
11. Hesse BW, Moser RP, Rutten LJ. Surveys of physicians and electronic health information. N Engl J Med 2010 Mar 04;362(9):859-860. [doi: 10.1056/NEJMc0909595] [Medline: 20200398]
12. Fox S. Mind the Gap: Peer-to-peer Healthcare. Pew Research Center. 2011 Aug 02. URL: https://www.pewresearch.org/internet/2011/08/02/mind-the-gap-peer-to-peer-healthcare/ [accessed 2021-11-13]
13. 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]
14. 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]
15. Why Most Facebook Users Get More than They Give. Pew Research Center. 2012 Feb 03. URL: https://www.pewresearch.org/internet/2012/02/03/why-most-facebook-users-get-more-than-they-give/ [accessed 2021-11-13]
16. Chiu Y, Hsieh Y. Communication online with fellow cancer patients: writing to be remembered, gain strength, and find survivors. J Health Psychol 2013 Dec;18(12):1572-1581. [doi: 10.1177/1359105312465915] [Medline: 23221492]

17. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. J Med Internet Res 2015 Jul 30;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
18. Thackeray R, Crookston BT, West JH. Correlates of health-related social media use among adults. J Med Internet Res 2013 Jan 30;15(1):e21 [FREE Full text] [doi: 10.2196/jmir.2297] [Medline: 23367505]
19. Bigley IP, Leonhardt JM. Extremity bias in user-generated content creation and consumption in social media. Journal of Interactive Advertising 2018 Sep 27;18(2):125-135. [doi: 10.1080/15252019.2018.1491813]
20. Kotliar DM. Depression narratives in blogs: a collaborative quest for coherence. Qual Health Res 2016 Jul 03;26(9):1203-1215 p. 1210, 1211. [doi: 10.1177/1049732315612715] [Medline: 26531881]
21. McGeehin Heilferty C. The search for balance: prolonged uncertainty in parent blogs of childhood cancer. J Fam Nurs 2018 May 06;24(2):250-270. [doi: 10.1177/1074840718772310] [Medline: 29732956]
22. Anderson JG, Hundt E, Dean M, Keim-Malpass J, Lopez RP. "The Church of Online Support": examining the use of blogs among family caregivers of persons with dementia. J Fam Nurs 2017 Feb;23(1):34-54. [doi: 10.1177/1074840716681289] [Medline: 27920340]
23. Andersson Y. Blogs and the art of dying: blogging with, and about, severe cancer in late modern Swedish society. Omega (Westport) 2019 Sep 12;79(4):394-413 p. 15. [doi: 10.1177/0030222817719806] [Medline: 28699836]
24. Heilferty C. Toward a theory of online communication in illness: concept analysis of illness blogs. J Adv Nurs 2009 Jul;65(7):1539-1547. [doi: 10.1111/j.1365-2648.2009.04996.x] [Medline: 19457009]
25. Høybye MT, Johansen C, Tjørnhøj-Thomsen T. Online interaction. Effects of storytelling in an internet breast cancer support group. Psychooncology 2005 Mar;14(3):211-220. [doi: 10.1002/pon.837] [Medline: 15386774]
26. Kraft MA. The use of blogs in medical libraries. Journal of Hospital Librarianship 2006 May;6(1):1-13. [doi: 10.1300/j186v06n01_01]
27. Sandars J. Twelve tips for using blogs and wikis in medical education. Med Teach 2006 Dec 03;28(8):680-682. [doi: 10.1080/01421590601106353] [Medline: 17594577]
28. Kim S, Chung DS. Characteristics of cancer blog users. J Med Libr Assoc 2007 Oct;95(4):445-450 [FREE Full text] [doi: 10.3163/1536-5050.95.4.445] [Medline: 17971894]
29. Huang L, Chou Y, Lin C. The influence of reading motives on the responses after reading blogs. Cyberpsychol Behav 2008 Jun;11(3):351-355 p. 351. [doi: 10.1089/cpb.2007.0063] [Medline: 18537507]
30. Rapp C. Aristotle's Rhetoric. The Stanford Encyclopedia of Philosophy. 2010. URL: https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/ [accessed 2021-11-13]
31. Cialdini R. The secret impact of social norms. RSA J 2007;154(5528):60. [doi: 10.1093/acprof:osobl/9780199743056.003.0007]
32. Cialdini RB. Harnessing the Science of Persuasion. Harvard Business Review. 2001 Oct. URL: https://hbr.org/2001/10/harnessing-the-science-of-persuasion [accessed 2021-11-13]
33. Resnik AJ, Cialdini RB. Journal of Marketing Research 1986 Aug;23(3):305. [doi: 10.2307/3151490]
34. de Wit JBF, Das E, Vet R. What works best: objective statistics or a personal testimonial? An assessment of the persuasive effects of different types of message evidence on risk perception. Health Psychol 2008 Jan;27(1):110-115. [doi: 10.1037/0278-6133.27.1.110] [Medline: 18230021]
35. liberman A, Chaiken S. Defensive processing of personally relevant health messages. Pers Soc Psychol Bull 2016 Jul 02;18(6):669-679. [doi: 10.1177/0146167292186002]
36. Slater MD, Rouner D. Value-affirmative and value-protective processing of alcohol education messages that include statistical evidence or anecdotes. Communication Research 2016 Jun 30;23(2):210-235. [doi: 10.1177/009365096023002003]
37. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. Psychol Bull 2001 Mar;127(2):267-286. [doi: 10.1037/0033-2909.127.2.267] [Medline: 11316014]
38. Yalch RF, Elmore-Yalch R. The effect of numbers on the route to persuasion. J Consum Res 1984 Jun;11(1):522-526 p. 526. [doi: 10.1086/208988]
39. Nexøe J, Kragstrup J, Søgaard J. Decision on influenza vaccination among the elderly. A questionnaire study based on the Health Belief Model and the Multidimensional Locus of Control Theory. Scand J Prim Health Care 1999 Jun 12;17(2):105-110. [doi: 10.1080/028134399750002737] [Medline: 10439494]
40. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
41. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
42. Weiss C. Mayo Clinic Q and A: The importance of a skin cancer check. Mayo Clinic. 2020 Mar 26. URL: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-q-and-a-the-importance-of-a-skin-cancer-check/ [accessed 2021-11-13]
43. Azjen I. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl J, Beckmann J, editors. Action Control. SSSP Springer Series in Social Psychology. Berlin, Heidelberg: Springer-Verlag; 1985:11-13.
44. Sheeran P, Conner M, Norman P. Can the theory of planned behavior explain patterns of health behavior change? Health Psychology 2001;20(1):12-19. [doi: 10.1037/0278-6133.20.1.12]

45. Conner M, Norman P, Bell R. The theory of planned behavior and healthy eating. Health Psychology 2002;21(2):194-201 [FREE Full text] [doi: 10.1037/0278-6133.21.2.194]
46. Armitage CJ. Can the theory of planned behavior predict the maintenance of physical activity? Health Psychol 2005 May;24(3):235-245. [doi: 10.1037/0278-6133.24.3.235] [Medline: 15898858]
47. Klein WMP, Stefanek ME. Cancer risk elicitation and communication: lessons from the psychology of risk perception. CA Cancer J Clin 2007 May 01;57(3):147-167 [FREE Full text] [doi: 10.3322/canjclin.57.3.147] [Medline: 17507441]
48. What Is Cancer? American Cancer Society. URL: https://www.cancer.org/treatment/understanding-your-diagnosis/what-is-cancer.html [accessed 2021-11-13]
49. Cialdini RB. Influence: The Psychology of Persuasion. New York, NY: Harper Business; 2007.
50. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. Am J Health Promot 1996 Aug 25;11(2):87-98. [doi: 10.4278/0890-1171-11.2.87] [Medline: 10163601]
51. Shulman R. Coronavirus patients turn to social media for answers, support. The Washington Post. 2020 May 23. URL: https://www.washingtonpost.com/health/coronavirus-patients-turn-to-social-media-group-for-answers-support/2020/05/22/cb715486-8fad-11ea-8df0-ee33c3f5b0d6_story.html [accessed 2021-11-13]
52. Nobles AL, Leas EC, Althouse BM, Dredze M, Longhurst CA, Smith DM, et al. Requests for diagnoses of sexually transmitted diseases on a social media platform. JAMA 2019 Nov 05;322(17):1712-1713 [FREE Full text] [doi: 10.1001/jama.2019.14390] [Medline: 31688875]

Abbreviations

ANOVA: analysis of variance
IRB: institutional review board
MTurk: Mechanical Turk

Edited by R Kukafka, G Eysenbach; submitted 04.08.20; peer-reviewed by E Neter, E Jablonski, A Efthymiou; comments to author 29.09.20; revised version received 17.11.20; accepted 15.10.21; published 22.12.21.

Please cite as:
Taylor KA, Humphrey Jr WF
Impact of Medical Blog Reading and Information Presentation on Readers' Preventative Health Intentions: Mixed Methods, Multistudy Investigation
J Med Internet Res 2021;23(12):e23210
URL: https://www.jmir.org/2021/12/e23210
doi:10.2196/23210
PMID:34941543

Original Paper

A Risk-Based Clinical Decision Support System for Patient-Specific Antimicrobial Therapy (iBiogram): Design and Retrospective Analysis

Lars Müller[1], PhD; Aditya Srinivasan[1], BA; Shira R Abeles[2], MD; Amutha Rajagopal[2], MD, MPH; Francesca J Torriani[2], MD; Eliah Aronoff-Spencer[1,2], MD, PhD

[1]Design Lab, University of California San Diego, La Jolla, CA, United States

[2]Division of Infectious Diseases and Global Public Health, Department of Medicine, UC San Diego Health, La Jolla, CA, United States

Corresponding Author:
Lars Müller, PhD
Design Lab
University of California San Diego
9500 Gilman Drive
MC0436 Atkinson Hall
La Jolla, CA, 92093
United States
Phone: 1 8582462639
Email: lmueller@tandemdiabetes.com

Abstract

Background: There is a pressing need for digital tools that can leverage big data to help clinicians select effective antibiotic treatments in the absence of timely susceptibility data. Clinical presentation and local epidemiology can inform therapy selection to balance the risk of antimicrobial resistance and patient risk. However, data and clinical expertise must be appropriately integrated into clinical workflows.

Objective: The aim of this study is to leverage available data in electronic health records, to develop a data-driven, user-centered, clinical decision support system to navigate patient safety and population health.

Methods: We analyzed 5 years of susceptibility testing (1,078,510 isolates) and patient data (30,761 patients) across a large academic medical center. After curating the data according to the Clinical and Laboratory Standards Institute guidelines, we analyzed and visualized the impact of risk factors on clinical outcomes. On the basis of this data-driven understanding, we developed a probabilistic algorithm that maps these data to individual cases and implemented iBiogram, a prototype digital empiric antimicrobial clinical decision support system, which we evaluated against actual prescribing outcomes.

Results: We determined patient-specific factors across syndromes and contexts and identified relevant local patterns of antimicrobial resistance by clinical syndrome. Mortality and length of stay differed significantly depending on these factors and could be used to generate heuristic targets for an acceptable risk of underprescription. Combined with the developed *remaining risk* algorithm, these factors can be used to inform clinicians' reasoning. A retrospective comparison of the iBiogram-suggested therapies versus the actual prescription by physicians showed similar performance for low-risk diseases such as urinary tract infections, whereas iBiogram recognized risk and recommended more appropriate coverage in high mortality conditions such as sepsis.

Conclusions: The application of such data-driven, patient-centered tools may guide empirical prescription for clinicians to balance morbidity and mortality with antimicrobial stewardship.

(J Med Internet Res 2021;23(12):e23571) doi:10.2196/23571

KEYWORDS

antimicrobial resistance; clinical decision support; antibiotic stewardship; data visualization

Introduction

Background

Antibiotic-resistant infections are widespread in the United States and across the globe and are an urgent threat to human health [1,2]. In the United States, over 2.8 million resistant infections occur each year, leading to delayed or ineffective therapy, longer hospital stays, and increased risk of death [1]. Patients who receive early empiric therapy that matches the in vitro susceptibility of their infection are up to 12 times more likely to survive (30-day crude mortality) [3], have shorter length of stays (LOSs) [4], and less long-term sequelae [5]. At the same time, overuse and prescription of overly broad antibiotics increases the risk of antimicrobial resistance (AMR), creating a positive feedback loop between behavior and ecological response, perpetuating a vexing health dilemma that spans sociobehavioral, ecological, and technical dimensions [6]. Increased antibiotic use in response to the COVID-19 pandemic exemplifies this complexity with the potential to further amplify the risk of AMR [7,8].

To combat the global AMR crisis today, leading agencies (eg, World Health Organization, Infectious Disease Society of America, Centers for Disease Control and Prevention, Agency for Healthcare Research and Quality, and the Clinical and Laboratory Standards Institute) have recommended that health care institutions produce and distribute locally derived antibiograms to clinicians, primarily as part of antimicrobial stewardship efforts [9,10]. Antibiograms have traditionally been high-level printed tabular summaries of local resistance patterns within an institution. These tables do not account for many factors known to influence antimicrobial susceptibility [11,12], even though such data are available in the electronic health record (EHR).

Recent machine learning reports demonstrated that EHR data can predict susceptibility, although generally for single disease types [13-15]. For instance, Yelin et al [13] used a data set of over 700,000 community-onset urinary tract infections (UTIs) to show that the use of prior antibiotics predicts AMR. The underlying machine learning techniques require large data sets and are therefore best suited for common infections, such as uncomplicated UTIs [13,15]. Different approaches are needed for smaller data sets, for example, to understand the influence of rare patient-specific factors or local susceptibility patterns.

Few hospitals have incorporated patient-specific and epidemiologic information into traditional antibiograms [16] or clinical decision support systems (CDSSs) [17-19]. Unit-specific or syndrome-specific antibiograms have been developed [20-22]. Visual analytics software have been studied as a tool for integrating data within an EHR on patient risk factors with isolate-specific susceptibilities [23]. The TREAT system is one of the most comprehensive and well-documented diagnostic and antimicrobial decision support tools [24,25]. TREAT uses a sophisticated causal probabilistic network [24] that hybridizes expert knowledge with heuristics and local antibiotic susceptibility data. In a multinational randomized controlled trial in 2013, TREAT use for inpatients led to a decreased LOS (1 day) but did not show mortality benefits when all variables were accounted for [25].

Reviews of clinical decision systems for antimicrobial management [17-19] found a paucity of evidence that these systems reduce mortality and morbidity when incorporated into daily clinical workflow, an unnatural segregation of outpatient and inpatient approaches, and a lack of systematic engagement with stakeholders about needs and workflow integration for support systems. A recent survey showed that only 44% of medical residents knew how to access the local antibiogram and preferred web-based resources such as UpToDate or the Sanford Guide [26]. These web-based treatment guidelines are not tailored to a specific health care environment as local antibiograms, but they are more accessible and provide explicit guidance for a case at hand. Thus, decades after the first digital innovations, the traditional antibiogram, with all its limitations, is still the most common AMR tool in use today.

Objectives

In this study, we introduce iBiogram as a digital CDSS for data-driven antimicrobial selection in the absence of definitive antibiotic susceptibilities. iBiogram addresses the limitations discussed above by examining infectious disease decision support as a complex sociotechnical problem [6,27]. Physicians and CDSS have to work together to complement the understanding of a specific case with guidance about the local epidemiology, likelihood of resistance, and associated risk of failure. Shared representations between algorithms and people are the key to blending the best qualities of expert knowledge and digital tools for efficient human technology teamwork in such cases [28]. Shared representations establish a common language to translate between clinical expertise and machine inference to enable efficient human-machine collaboration. To this end, we explored diverse patient information and antimicrobial testing results, presented new visualizations of insights generated from these data, and share the results of evaluating a new prototype and risk-based metric compared with historic provider behavior.

Methods

Overview

We systematically evaluated a comprehensive clinical data set that crossed ambulatory and hospital settings and contained antimicrobial testing, clinical context, and patient factors. We then identified local factors influencing microbial prevalence and AMR, their impact on mortality, and how they combine to influence the success of antibiotic treatments. Finally, we developed and tested a decision support tool that may be used in both low-risk ambulatory settings where overprescriptions likely dominate and inpatient settings where the risks and benefits of early and accurate empiric therapy are the greatest.

Data Source and Analysis

We analyzed 5 years of antibiotic susceptibility isolate testing data and related deidentified patient information from the University of California San Diego Health Sciences (UCSDHS). The study was reviewed and approved by the University of California San Diego Human Research Protections Program

(Institution Review Board #161853). Positive bacterial cultures with susceptibility results between May 2011 and November 2016 were included. Contents included identified pathogens; their susceptibility to tested antibiotics; diagnostic information from before, during, and after the ordering of cultures; any medications prescribed before and throughout their visit; the problem list at the time of order; and general demographic information.

We processed the data before analysis by removing repeat susceptibility tests according to the Clinical and Laboratory Standards Institute guidelines [9]. Suppressed antibiotics and supplemental tests performed after cultures returned were not considered for the purposes of this study. The International Classification of Diseases-10 (ICD-10) codes at order and at discharge were mapped to syndromes and prevalent comorbidities; for example, sepsis was defined as a positive blood culture and a corresponding ICD-10 code. An overview of the mapping is provided in Multimedia Appendix 1.

We studied the influence of time, demographics, assigned syndromes, medications, and comorbidities on the resulting distribution of causal bacteria, their susceptibility, and clinical outcomes. Logistic regression was used to calculate the odds ratios (OR) adjusted for age and sex, and the results were filtered by significance ($P<.05$). The influence of ineffective therapy on clinical outcomes was assessed between syndrome-factor combinations. Critical combinations were identified by evaluating changes in 7-day all-cause mortality and median LOS for prescribed treatments.

The impact of factors and comorbidities on the resulting distribution of causal bacteria, their susceptibility, and mortality were analyzed using the Python statsmodels package implementation of logistic regression adjusted for age and sex to account for significant differences in syndrome types and commonly occurring organisms, for example, higher rates of UTI and gram-negative bacteria in women. Age was modeled as a categorical variable (0-14, 15-24, 25-44, 45-64, and ≥65). The impact of age and sex was evaluated by controlling for each other. The change in LOS was analyzed using the Mann-Whitney *U* test from the scipy package. The effect size in days was calculated as the difference in medians to account for skewed distribution.

Algorithm and iBiogram Decision Support Platform Development

To analyze and communicate the impact of risk factors for specific cases, we developed the remaining risk metric as a shared representation [28]. The risk of ineffective empiric antibiotic therapy was calculated as the sum of the individual resistance probabilities of each possible causal agent r_i weighted by the probability of that causal agent a_i. With n referring to the number of possible causal agents:

The remaining risk of combination therapy was calculated by selecting the minimal r_i of all therapy antibiotics. The result was a prevalence-weighted multiantibiotic antibiogram that we refer to as iBiogram. We then developed a web-based prototype of the iBiogram where all statistically significant factors may be coselected to generate a patient-centric antibiogram, as well as a recommended set of antibiotics with the best coverage for predicted organisms.

The iBiogram algorithm draws antibiotics from a knowledge base that maps antibiotics to syndromes. Antibiotics are additionally filtered by the available test data for the selected scenario to prevent inflated rankings caused by a few or, in the worst case, a single antibiotic susceptibility isolate test. All possible antibiotic combinations were generated and ranked based on the remaining subset of antibiotics. Sepsis is currently the only syndrome that allows the combination of three antibiotics with amikacin and gentamicin reserved as the third drug.

Evaluation

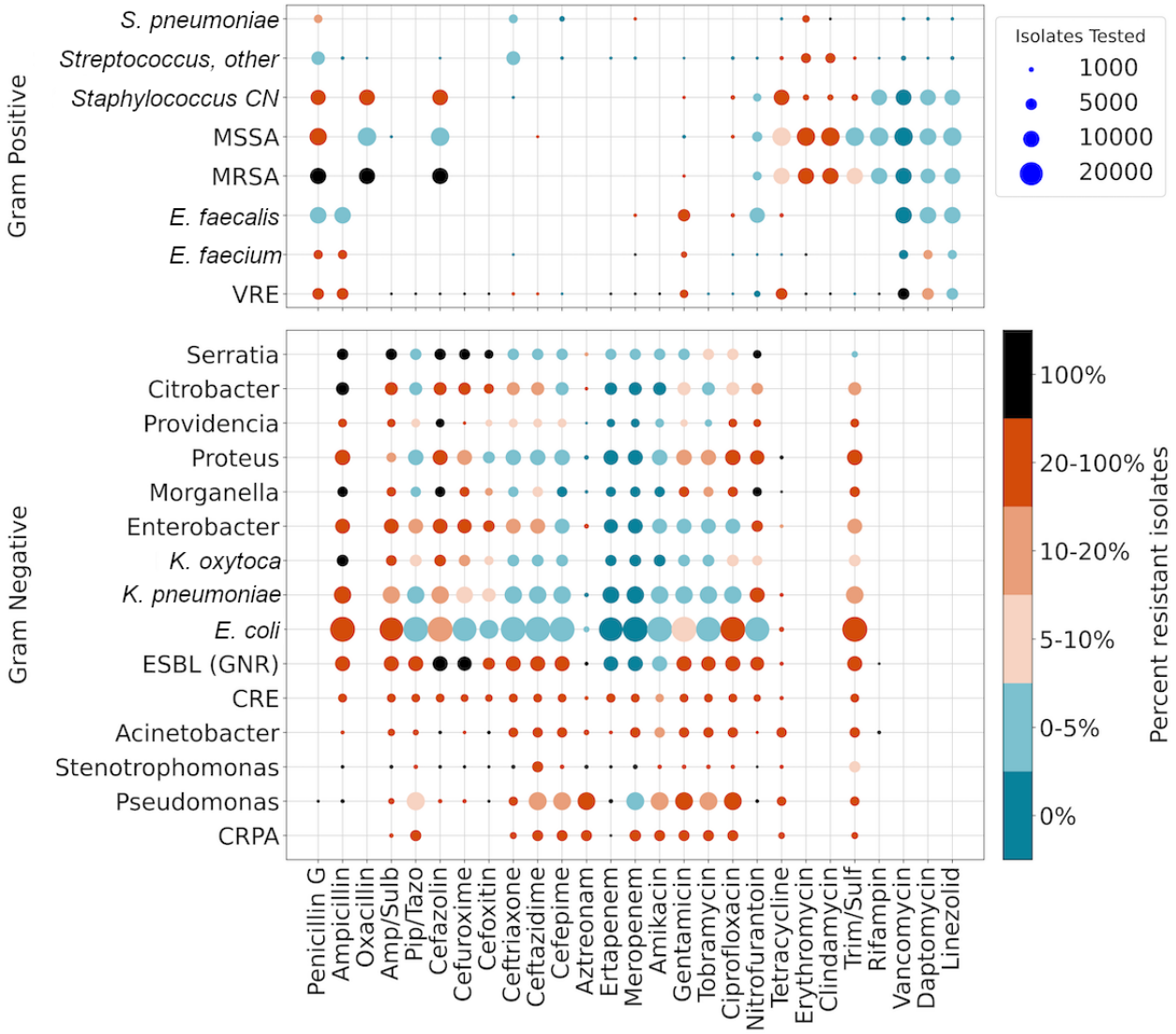
The iBiogram algorithm was evaluated by using data before the last 6 months and comparing the performance of predicted *best* treatments to the actual treatments and outcomes in the last 6 months of the data. Physicians' decisions and iBiogram prediction decisions were evaluated as successful if the identified organism was susceptible to any of the prescribed or recommended antibiotics. For physicians' empiric therapy choices, we considered antibiotics prescribed between 24 hours before and 8 hours after the culture was ordered. Considering antibiotic stewardship and building on the conducted mortality analysis, we used predicted-susceptibility targets developed by Cressman et al [29] to rank potential treatments for a given syndrome based on the risk of mortality or prolonged LOS. Prescribed and suggested therapeutic regimens were further scored on the use of protected antibiotics and essential medicines according to World Health Organization's Action for Welfare and Awakening in Rural Environment Classification Database [30]. This classification groups therapies into *Access group Antibiotics*, those that are broadly active but minimal risk for resistance, *Watch group antibiotics* that are high priority therapies at substantial risk for developing resistance without stewardship, and *Reserve group antibiotics* that are saved only for use in life-threatening multidrug-resistant infections.

Results

Data Summary

Figure 1 depicts an overview of all susceptibility tests for gram-positive and gram-negative organisms over the full 5-year data set at the UCSDHS. A complete table is available in the Multimedia Appendix 2. Here, we re-envision a general static antibiogram, summarizing prevalence (dot area) and resistance patterns (red is increasing prevalence). Although printed antibiograms display only the current quarter, including more data can result in a higher predictive value for a specified patient, as a larger number of cases reduces the effect of outliers. An analysis of larger time windows found no significant differences between quarterly antibiograms and data, including up to 7 quarters (Multimedia Appendix 3).

Figure 1. Overview of antibiotic resistance: Summary of the analyzed data for the most commonly tested antibiotics and causal agents. Each dot represents all isolates with a given causal agent (horizontal) and tested antibiotic (vertical). The number of isolates is indicated by dot-size and the probability of resistance is encoded as color ranging from blue (0% resistance) over red (>20% resistance) to black (100% resistance). CN: Coagulase Negative Staphylococci; CRE: Carbapenem-resistant enterobacteriaceae; CRPA: Carbapenem-resistant Pseudomonas aeruginosa; ESBL: Extended Spectrum Beta Lactamase producing Enterobacteriaceae; GNR: Gram negative rod; MRSA: Methicillin-resistant Staphylococcus aureus; MSSA: Methicillin-susceptible Staphylococcus aureus; VRE: Vancomycin-resistant Enterococcus.



Factors Influencing Microbial Ecology, Resistance, Mortality, and LOS

The risk factors influencing organism prevalence and resistance at UCSDHS range from well-known contributing factors such as prior antibiotic use to more specific insights, such as the influence of a specific transplant history. Figure 2 depicts the principal factors that drive AMR, Multimedia Appendices 3-6 supply additional data. They were assessed and reported as adjusted ORs for changes in organism prevalence (left) and resistance (right). Expected contributing factors included sex, age, and context, such as whether a patient presented in an outpatient setting, to an emergency department, or developed illness while hospitalized. Certain historic patient factors (eg, diabetes, cystic fibrosis, hemodialysis, or history of transplantation) were linked to significantly higher rates of multidrug-resistant organisms. For instance, patients on hemodialysis had higher rates of carbapenem-resistant Enterobacteriaceae (CRE; OR 2.4, 95% CI 1.7-3.3). Patients with health care–associated pneumonia were at increased risk for *Acinetobacter* (OR 3.4, 95% CI 2.7-4.4), CRE (OR 3.3, 95% CI 2.7-4.1), and carbapenem-resistant *Pseudomonas aeruginosa* (CRPA; OR 3.3, 95% CI 2.7-4.1). Cystic fibrosis presented the strongest effects in both ecology and resistance, showing drastic shifts to *Pseudomonas* (OR 12.4, 95% CI 11.4-13.5), CRPA (OR 22.7, 95% CI 18.8-27.3), *Stenotrophomonas* (OR 5.5, 95% CI 4.5-6.7), and methicillin-resistant *Staphylococcus aureus* (OR 1.7, 95% CI 1.5-1.9), with significant decreases in other pathogenic organisms.

The syndrome-associated 7-day all-cause mortality, as well as median LOS, differed significantly based on sensitivity to empiric treatments. The specific syndrome factors are reported in Figure 3. Failure to prescribe active antibiotics for patients presenting with a UTI or community-acquired pneumonia did not significantly influence mortality or resulted in small changes

in the LOS, whereas cases of health care–associated pneumonia and sepsis showed significant increases in mortality and LOS. For instance, nosocomial sepsis with hematologic malignancy was significantly associated with increased 7-day all-cause mortality (OR 3.2) and LOS (+19 days) when empiric antibiotics were inactive against the causative pathogen.

Figure 2. Factors influencing microbial prevalence (left) and resistance (right): Odds ratios for statistically significant factors (P<.05) influencing the probability of encountering a specific bacterium or an isolate resistant to a particular treatment were adjusted by sex and age. Sex and age were adjusted against each other. Factors include demographic information, zip code, medications, syndromes, and comorbidities as coded in International Classification of Diseases-10. In the left graph, purple indicates a higher prevalence of particular pathogens. In the right graph, red indicates a higher and blue a lower prevalence of resistant isolates. CAP: community-acquired pneumonia; CN: coagulase-negative Staphylococci; COPD: chronic obstructive pulmonary disease; CRE: carbapenem-resistant Enterobacteriaceae; CRPA: carbapenem-resistant Pseudomonas aeruginosa; ESBL: extended-spectrum β-lactamases; GNR: Gram negative rod; HAP: health care–associated pneumonia; ICU: intensive care unit; MRSA: methicillin-resistant Staphylococcus aureus; MSSA: methicillin-susceptible Staphylococcus aureus; SSTI: skin and soft tissue infection; UTI: urinary tract infection; VRE: vancomycin-resistant Enterococcus.

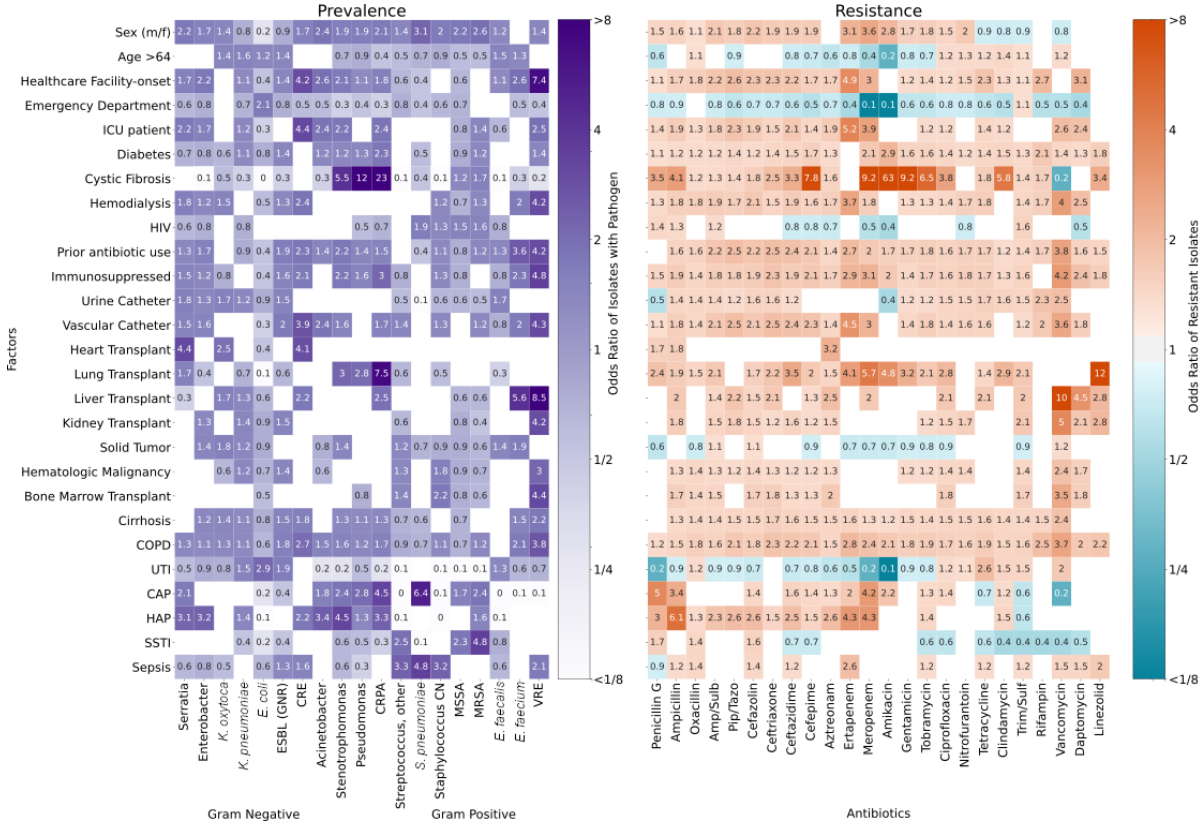
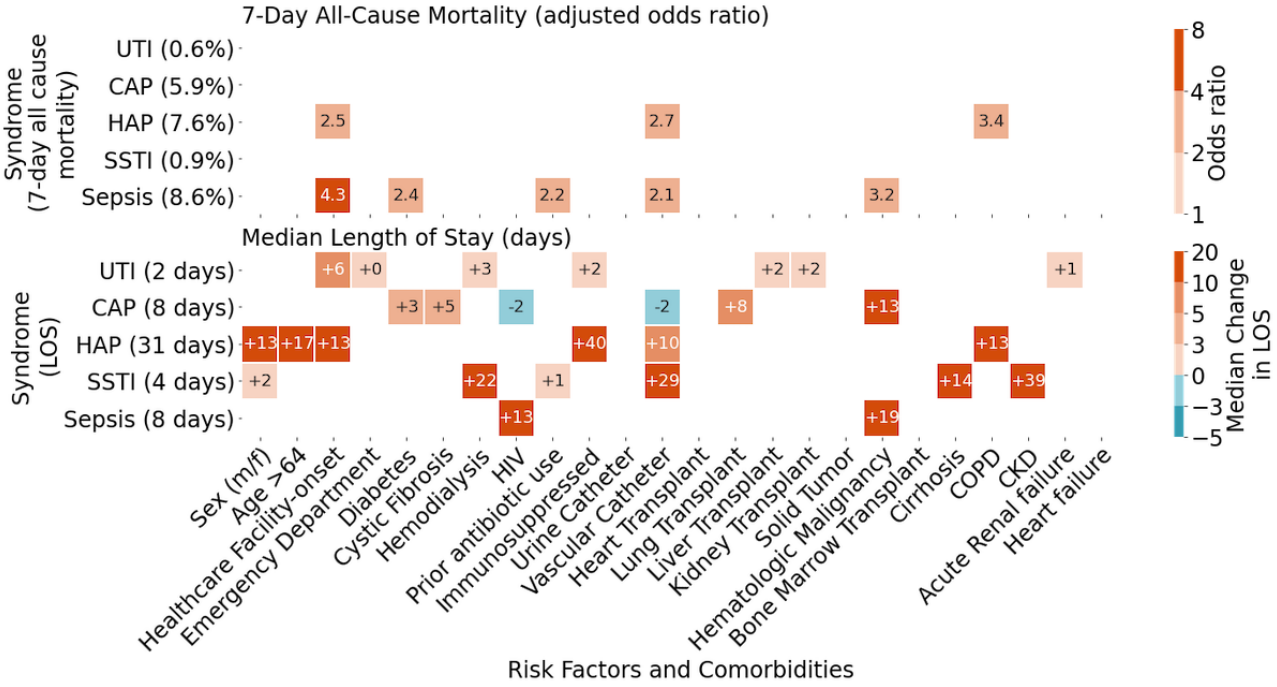


Figure 3. Influence of treatment failure on 7-day all-cause mortality (top) and median length of stay (bottom) in syndromes/risk factor combinations: A summary of the influence of resistant pathogens for syndromes modified by risk factor or comorbidity. Only significant results (P<.05) are shown. The top graph shows the age and sex adjusted 7-day all-cause mortality odds ratio for failed treatments, for example, hospital-facility sepsis patient with ineffective treatment are 4.3 times more likely to die within 7 days. The bottom graph depicts how ineffective treatment affects the median length of stay for the same combinations including only surviving patients and testing significance using the Mann-Whitney U test. For each syndrome, the mean mortality and the median length of stay is listed on the y-axis behind the syndrome name in brackets. CAP: community-acquired pneumonia; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; HAP: health care–associated pneumonia; SSTI: skin and soft tissue infection; LOS: length of stay; UTI: urinary tract infection.

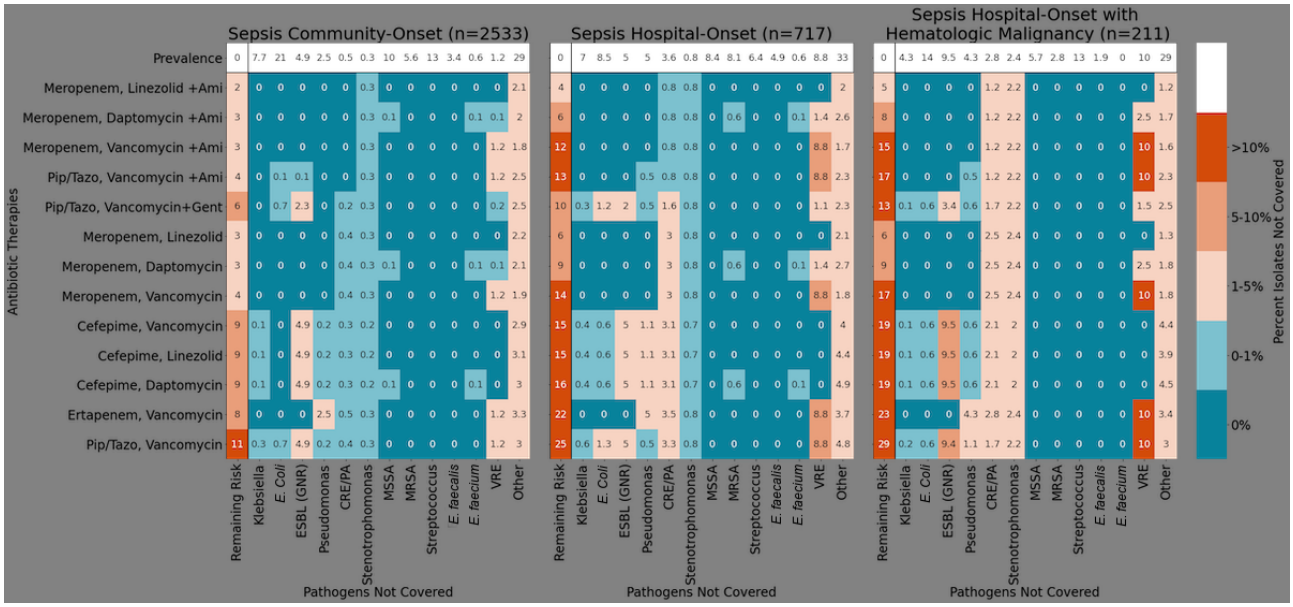


Visualizing Risk of Resistance for Empiric Therapy

The difficult tradeoff between overly broad empiric regimens and the risk of treatment failure requires a shared representation that accounts for known risk factors. Consider the scenario shown in Figure 4 where on the left, we show a filtered remaining risk antibiogram displaying the performance (chance of failure) of common empiric antibiotic regimens for patients presenting with community-onset sepsis in the emergency department. In the middle, we consider changes in risk associated with hospital-onset sepsis and on the right for the subset of these patients with a hematologic malignancy.

An antibiogram for community-onset sepsis would predict the frequently prescribed empiric therapy of vancomycin and piperacillin/tazobactam to fail in 11% of pathogens in this data set. However, after adding hospital-onset and hematologic malignancy factors, the failure risk increases to 25% and over 28%, respectively, each below the recommended target of complete or at a minimum 90% coverage (<10% remaining risk) in cases of severe sepsis [29]. At first glance, daptomycin and meropenem, less traditional first-line agents for hospital-onset sepsis, might be considered a better empiric regimen in these populations, with only 9% remaining risk of failure.

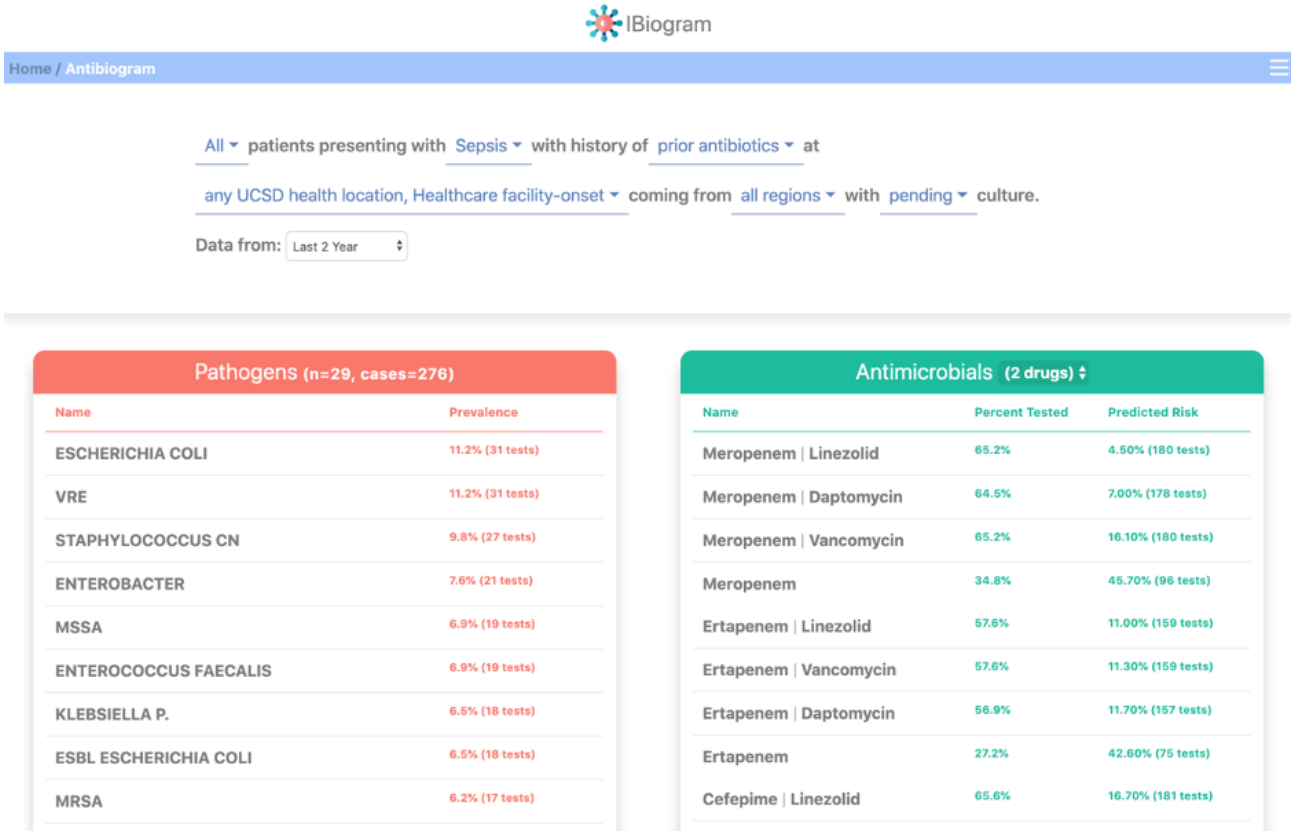
Figure 4. Remaining risk of treatment failure in sepsis: A summary of the percent risk of failure for common multidrug regimens in sepsis cases. Shown are blood culture summations from community-onset sepsis, hospital-onset sepsis, and hospital-onset sepsis in patients with a hematologic malignancy. Numeric values represent remaining risk (prevalence x predicted resistance) to a given antibiotic or antibiotics. The prevalence (top row) lists the probability of encountering the delineated causal agents. If a treatment does not cover a pathogen the risk equals the prevalence of that pathogen, for example, meropenem never covers methicillin-resistant Staphylococcus aureus. The percentage in the remaining risk column (leftmost) sums the probability of encountering a pathogen not covered by the delineated therapy. Blue color coding represents <1%, increasing in red intensity until, red ≥10%, a minimum threshold suggested for severe infection [29]. CRE: carbapenem-resistant Enterobacterales; ESBL: extended-spectrum β-lactamases; GNR: Gram negative rod; MRSA: methicillin-resistant Staphylococcus aureus; MSSA: methicillin-susceptible Staphylococcus aureus; PA: Pseudomonas aeruginosa; VRE: vancomycin-resistant Enterococcus.



Envisioning a Human-Centered Antibiogram

In Figure 5, we provide a prototype digital antibiogram capable of accounting for the factors that predict the particular ecology associated with patients sharing these factors, along with the associated resistance patterns and suggested antibiotics with their expected coverage. Here, we selected all patients over the last eight quarters who developed nosocomial sepsis while receiving antibiotic therapy. In the left panel, we can see the most prevalent pathogens in this population. The user can scroll down to see the long tail of the organisms. On the right, a list of potential treatments (1-3 drugs may be displayed) is shown with predicted coverage, and the ability to toggle and show the remaining risk. Overall, using a combination of patient factors, historic culture data, and antibiotic prescribing rules, the user can readily select key factors to generate a patient-level antibiogram that is tailored to the case at hand. The system then provides a list of therapeutic options (ranked either by predicted coverage or by remaining risk) and suggests treatments that a provider may select based on further details such as patient history of medication allergies, need for bactericidal versus bacteriostatic drugs, preferred route of administration, drug-drug interactions, and cost of treatment.

Figure 5. The iBiogram digital antibiogram: iBiogram is a web-based app that allows the user to select patient demographics, presenting syndrome, prior use of antibiotics or immune suppression, comorbid conditions, as well as the location and context of infection, the region a patient has come from and if the pending cultures have been identified as gram-positive or -negative. In the example above, all patients with diabetes on prior antibiotics and presenting with nosocomial sepsis over 2 years are shown. If the user suspects a specific organism based on history that organism can be selected, and the table will reorient to show the sensitivities for just that organism. Likewise, a user can select a treatment regimen and examine the organisms that are sensitive or resistant to the treatment. ESBL: extended-spectrum β-lactamases MRSA: methicillin-resistant Staphylococcus aureus; MSSA: methicillin-sensitive susceptible Staphylococcus aureus; VRE: vancomycin-resistant Enterococcus.



Comparing Empiric Therapy at the Time of Order Versus iBiogram Suggestions

Given that patients who received effective empiric therapy were more likely to survive and had shorter LOS, we analyzed the failure rate of prescribed empiric therapies based on final sensitivity results and compared them to the suggested iBiogram regimens (Table 1). For patients presenting to the emergency department with a UTI, the most commonly prescribed antibiotics were ciprofloxacin and ceftriaxone, and the overall failure rate was 17.7% (85/479), whereas the suggested iBiogram treatments would have been between 23% and 15.1%. For the subset of patients with UTIs in the emergency department with prior antibiotics, the empiric therapy treats only 74% (48/65; 26.2% failure), whereas suggested regimens would have achieved a reduction of 3.6%-14.2%. For community-onset sepsis, the most frequently prescribed drugs were piperacillin/tazobactam and vancomycin, and the overall failure rate considering all cases was 8.1% (31/383), compared with 5.5% for the iBiogram-suggested treatment. For cases of hospital-onset sepsis, however, where there is significant associated mortality, prescribed empiric failure rose to 16% (12/75) in comparison to 7.8% using the proposed antibiotics. Finally, in cases of hospital-onset sepsis in patients with hematologic malignancy, where we found the strongest correlation of mortality and LOS with sensitivity to prescribed antibiotics, we see failure of empiric therapy climb to 17.4% (4/27) failure, whereas suggested combinations would have failed in only approximately 13% of cases.

Table 1. Comparison of the performance of iBiogram recommendations to actual antibiotics prescribed by doctors at the time of order.[a]

Syndrome	Risk factor		Received empiric antibiotics		Retrospective comparison of prescribed empiric therapies to iBiogram suggestions				
	Total	Deceased, n (%)	Total	Deceased, n (%)	Empiric therapy (%)				iBiogram suggestion
					Measured risk	AWaRe[b] group access	AWaRe group watch	AwaRe group reserve	Measured risk (%); predicted risk; AWaRe group
UTI[c] in the ED[d]	939	0 (0)	479	0 (0)	17.7	24	76	0	• 23%, nitrofurantoin; 19.4%; access • 15.1% ceftriaxone; 17.8%; watch
UTI in the ED on antibiotics	147	0 (0)	65	0 (0)	26.2	26	74	0	• 1.9%, amikacin; 19.3%; access • 18.7% nitrofurantoin and trimethoprim/sulfamethoxazole; 19.9%; access
Sepsis, community onset	562	43 (7.7)	383	30 (7.8)	8.1	6	92	2	• 14.2%, vancomycin and piperacillin/tazobactam; 9.8%; watch • 5.5% vancomycin and ertapenem; 8.2%; watch
Sepsis, hospital onset	181	17 (9.4)	75	7 (9.3)	16	1	95	4	• 7.8% daptomycin and meropenem; 7.6%; reserve • 3.6% daptomycin and meropenem+amikacin; 6.8%; reserve
Sepsis, hospital onset with hematologic malignancy	39	8 (20.5)	23	4 (17.4)	17.4	4	91	4	• 15.5% daptomycin and meropenem; 7.4%; reserve • 12.9% daptomycin and meropenem+gentamicin; 5.1%; reserve

[a]Compares observed prescription and success rates for prescribers in the last 6 months of the study versus suggested iBiogram regimens for cases with prescriptions based on all prior data. Noted mortality rates are a proxy for risk of using ineffective antimicrobial therapy.

[b]AWaRe: Action for Welfare and Awakening in Rural Environment.

[c]UTI: urinary tract infection.

[d]ED: emergency department.

Discussion

Research in Context

On the basis of gaps in the science noted by Curtis [17], Rawson [18], and Laka [19], this study adopted a systematic human-centered approach to analyze, visualize, and operationalize factors associated with AMR, mortality, and LOS of bacterial infections across a United States academic health care system, spanning from ambulatory and emergency to inpatient and intensive care settings. Our efforts differed from machine learning approaches that remain limited to large homogenous data sets, as well as from the robust TREAT program [25], which incorporates more assumed knowledge but may miss factor associations. This study mirrors significant efforts such as Detecting and Eliminating Bacteria Using Information Technologies (DeBUGIT) [31], Wise Antimicrobial Stewardship Support System (WASPSS) [32], and EPiC IMPOC (Enhanced, Personalized and Integrated Care for Infection Management at the Point-Of-Care) [33,34] that have each sought to optimize bug-drug combinations by fusing data-driven approaches with expert consensus or guidelines. However, we propose that our integration of morbidity and mortality with clinical and microbiological data to present the risk of failure in a mortality cost-benefit context may be more intuitive to clinicians balancing patient safety and antimicrobial stewardship across the continuum of care.

We conducted a comprehensive examination of the factors influencing microbial ecology and resistance within a large urban and suburban health care system to set the stage for the development of an antibiotic recommendation system that can support clinicians prescribing antibiotics in the absence of culture data. We identified predictive factors that may be used to create tailored empiric antibiotic choices; however, these factors may be difficult to convey using static and nondigital approaches. Alone, these factors are also insufficient to make optimal therapeutic decisions, as recommendations to prescribers

require fusion of purely probabilistic approaches with clinical heuristics, expert guidelines, and other risk benefit considerations [24].

We cleaned and created a new graphical representation of the bacterial prevalence and antibiotic resistance data. At a glance, this approach can be seen to be more readily informative of both the likelihood of encountering given pathogens and their expected pattern of resistance than current paper-based antibiograms. This approach also demonstrates the limitations of a hospital-wide and static view of AMR when attempting to map data to a specific patient and immediately suggests that more granular patient-specific data could inform better decision-making. Such a graphical overview could be generated for subsets of the data to generate and track a *fingerprint* of resistance patterns.

The complex sepsis cases in Figure 4 highlight the tension between the need for more precise, patient-specific, and readily accessible antibiograms but also indicate that actionable support for clinicians requires a complex understanding of patient factors and best practices that may be specific to an institution. For instance, closer examination and expert clinical experience revealed that daptomycin may offer little advantage over vancomycin against methicillin-resistant *Staphylococcus aureus*, merely adding coverage against vancomycin-resistant enterococci, which is generally encountered in specific situations and may be associated with higher mortality if missed empirically in these cases [35]. Likewise, the benefits of using a third agent such as amikacin, a potentially ototoxic and nephrotoxic drug, over the safer meropenem and daptomycin combination, would only be worthwhile in specific risk groups with higher rates of CRE and CRPA, such as patients with cystic fibrosis or a history of lung transplantation. The robust performance of amikacin in our data set also highlights the potential influence of a hospital's formulary, as amikacin is rarely used and is not routinely stocked in the study hospital, where other aminoglycosides are favored. The ability to track antibiotic use with antibiotic resistance data on broad scales and among various institutions with different prescribing practices will ultimately shed important light on antimicrobial stewardship priorities.

To this end, we developed a digital tool combining up-to-date hospital data with clinical rules (here syndrome-specific antibiotics from expert guidelines) as an interactive precision tool for the selection of antibiotic combinations. This tool covers many typical scenarios encountered by clinicians across the continuum of care, ranging from empiric treatment of a UTI in ambulatory patients to solid-organ transplant recipients with sepsis presenting to the emergency room and the care of complex hospitalized patients such as those with stem cell transplants or cystic fibrosis. Notably, this ecology is highly regional and often hospital specific, highlighting the need to incorporate local organism prevalence, local formulary, and patient-specific risk factors. In the future, the tool could be readily integrated into EHRs to alert clinicians when prescribed regimens are too broad in coverage or have higher than accepted chances of failure. Alternatively, more appropriate regimens can be suggested. Eventually, such tools could be used for individual patients with complicated infectious disease histories or as personalized antibiograms in ambulatory settings. This might benefit those with frequent infections such as UTI, dialysis patients with repeated sepsis, or cystic fibrosis patients with highly specific ecologies.

Our data underscore the potential benefits of more specific antibiograms as well as the continued need for careful, case-based consideration of therapeutic options by expert clinicians. Although the data presented here support the utility of multidimensional antibiograms, they also highlight the challenges in designing a balanced tool that offers generalizability to multiple syndromes and presents therapeutic options that account for the immediate need for broad coverage and the counterpoised threat of antibiotic resistance in the clinical environment. Combining associated mortality data that is specific to certain patient populations may help navigate this risk so that narrow-spectrum antibiotic regimens may be used in low-risk scenarios and broader spectrum treatments could be saved for short periods in high-stakes scenarios. Table 1 illustrates this point. Here, we see that iBiogram not only distinguished between syndromes with no mortality (UTI) and those with high (sepsis), similar to providers, but tailored treatments further based on other risk factors. Notably, providers prescribed 74% (48/65) to 75.9% (364/479) of watch antibiotics in UTI, whereas iBiogram recommended watch group therapies only in UTI cases with prior antibiotic use. In sepsis, 91% (21/23) to 95% (71/75) of provider prescriptions were the watch group and 2.1% (8/383) to 4% (3/75) reserve antibiotics, whereas the iBiogram system suggested reserve antibiotics only in high mortality cases, providing better empiric coverage.

Limitations and Future Directions

The retrospective nature of the study and the quality of the EHR data are the two primary limitations. Retrospectively distinguishing pathogenic organisms from contaminants can be challenging. Therefore, we excluded many positive cultures that did not have a corresponding ICD-10 code. Moreover, many infections are treated without culture or negative cultures are ignored. This leads to a selection bias in the analyzed EHR data, whereby only positive cultures are recorded, likely resulting in a bias toward higher acuity diseases. The presented factors rely mainly on ICD-10 codes that (1) limit granularity, (2) are prone to coding errors, and (3) may not reflect important uncharted factors, including severity and timing of disease. For instance, it was not possible to distinguish sepsis from severe sepsis based on coding alone and medical data such as blood pressure, or use of pressers, were beyond the scope of this work. Furthermore, not all antibiotics were tested against all the isolates. For example, levofloxacin is rarely tested at our institution because its resistance patterns closely match those of ciprofloxacin. Finally, new antibiotics are continuously brought to the market to treat resistant organisms, and retrospective data will, by definition, not include the newest available antimicrobials.

Similar to traditional antibiograms, the prototype described in this work currently focuses only on antibiotic coverage and does not consider other factors such as side effects, or restricted institutional antibiotics, nor does it distinguish the mode of delivery (intravenous vs oral) and the cost or risk of interactions

with other medications. It also did not present patient-specific data with regard to recent specific antibiotic exposure or past personal culture data. Finally, the current version of the iBiogram did not include data on newer broad-spectrum antimicrobials such as ceftolozane-tazobactam and ceftazidime-avibactam, which may offer more reasonable and less toxic options in the setting of a CRE or CRPA.

Finally, we propose that these methods may be applied to inter–health care system data to create regional antibiograms [36-38] or with international databases such as WHONET data, where the platform could also extend to resource-limited and rural settings [39]. Therefore, data sets grow in breadth, depth, and history, machine learning tools and recursive approaches may increasingly be applied to predict the drivers of ecology and patterns of resistance.

Conclusions

Overall, we aimed to create an interactive shared representation of the complex factors that clinicians must navigate for effective empiric prescriptions. To this end, we developed a CDSS that strives to promote effective collaboration with users, bridging the gap in presentation and reasoning between clinical guidelines, clinicians' expertise, and data science by representing local outcomes and local ecology in a user-centered tool. Overall, we propose that such tools could improve clinical and safety outcomes, reduce adverse events because of inappropriate antibiotics, provide real-time suggestions to prescribers, and lower AMR and costs of care.

Acknowledgments

Data extraction was partially supported by the National Institutes of Health (grant UL1TR001442 of Clinical and Translational Science Awards program. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions and International Classification of Diseases-10 (ICD-10) mapping. Overview of definitions used for factors, including the mapping of ICD-10 values to syndromes and comorbidities.
[DOC File , 88 KB - jmir_v23i12e23571_app1.doc]

Multimedia Appendix 2
Overview of cohort and culture properties after removing duplicate cultures showing the number of patients, visits, isolates, and susceptibility tests for each factor.
[DOC File , 118 KB - jmir_v23i12e23571_app2.doc]

Multimedia Appendix 3
Infection ecology over time demonstrates the ecology of infection-causing bacteria in the hospital system. Printed antibiograms generally consider only the past quarter of the hospital data. Here, we analyze the past quarter's ecology compared with the past two quarters, past three quarters, etc. Overlaid represents the *P* value of each comparison. As shown, the bacterial ecology when considering the past four quarters is statistically similar ($P>.05$) to that of the past quarter, suggesting that we can use up to four quarters).
[PNG File , 64 KB - jmir_v23i12e23571_app3.png]

Multimedia Appendix 4
All factors influencing microbial prevalence (left) and resistance (right). Odds ratios for statistically significant factors ($P<.05$) influencing the probability of encountering a specific bacterium or an isolate resistant to a particular treatment were adjusted by sex and age. Sex and age were adjusted for each other. Factors included demographic information, zip code, medications, syndromes, and comorbidities as coded in the International Classification of Diseases-10. In the left graph, purple indicates a higher prevalence of pathogens. In the right graph, red indicates a higher and blue indicates a lower prevalence of resistant isolates.
[PNG File , 692 KB - jmir_v23i12e23571_app4.png]

Multimedia Appendix 5
Patient home address influencing microbial prevalence (left) and resistance (right). Odds ratios for statistically significant factors ($P<.05$) influencing the probability of encountering a specific bacterium or an isolate resistant to a particular treatment were adjusted by sex and age. Factors included demographic information, zip code, medications, syndromes, and comorbidities as coded in the ICD-10. In the left graph, purple indicates a higher prevalence of pathogens. In the right graph, red indicates a higher and blue indicates a lower prevalence of resistant isolates.

[PNG File , 147 KB - jmir_v23i12e23571_app5.png]

Multimedia Appendix 6
Factors influencing microbial prevalence (left) and resistance (right) in sepsis Odds ratios for statistically significant factors (P<.05) influencing the probability of encountering a specific bacterium or an isolate resistant to a particular treatment were adjusted by sex and age. Factors included demographic information, zip code, medications, syndromes, and comorbidities as coded in the International Classification of Diseases-10. In the left graph, purple indicates a higher prevalence of pathogens. In the right graph, red indicates a higher and blue indicates a lower prevalence of resistant isolates.
[PNG File , 447 KB - jmir_v23i12e23571_app6.png]

References

1. Antibiotic resistance threats in the United States, 2019. Centers for Disease Control and Prevention. URL: https://stacks.cdc.gov/view/cdc/82532 [accessed 2021-08-18]
2. -. The antibiotic alarm. Nature 2013 Mar 14;495(7440):141. [doi: 10.1038/495141a] [Medline: 23495392]
3. Lee C, Lee C, Yang C, Hsieh C, Tang H, Ko W. Beneficial effects of early empirical administration of appropriate antimicrobials on survival and defervescence in adults with community-onset bacteremia. Crit Care 2019 Nov 20;23(1):363 [FREE Full text] [doi: 10.1186/s13054-019-2632-1] [Medline: 31747950]
4. Fraser A, Paul M, Almanasreh N, Tacconelli E, Frank U, Cauda R, TREAT Study Group. Benefit of appropriate empirical antibiotic treatment: thirty-day mortality and duration of hospital stay. Am J Med 2006 Nov;119(11):970-976. [doi: 10.1016/j.amjmed.2006.03.034] [Medline: 17071166]
5. Scarsi KK, Feinglass JM, Scheetz MH, Postelnick MJ, Bolon MK, Noskin GA. Impact of inactive empiric antimicrobial therapy on inpatient mortality and length of stay. Antimicrob Agents Chemother 2006 Oct;50(10):3355-3360 [FREE Full text] [doi: 10.1128/AAC.00466-06] [Medline: 17005817]
6. Malik B, Bhattacharyya S. Antibiotic drug-resistance as a complex system driven by socio-economic growth and antibiotic misuse. Sci Rep 2019 Jul 05;9(1):9788 [FREE Full text] [doi: 10.1038/s41598-019-46078-y] [Medline: 31278344]
7. Hsu J. How covid-19 is accelerating the threat of antimicrobial resistance. Br Med J 2020 May 18;369:m1983. [doi: 10.1136/bmj.m1983] [Medline: 32423901]
8. Record number of countries contribute data revealing disturbing rates of antimicrobial resistance internet. World Health Organization. URL: https://www.who.int/news-room/detail/01-06-2020-record-number-of-countries-contribute-data-revealing-disturbing-rates-of-antimicrobial-resistance [accessed 2021-08-18]
9. Hindler J, Stelling J. Analysis and presentation of cumulative antibiograms: a new consensus guideline from the Clinical and Laboratory Standards Institute. Clin Infect Dis 2007 Mar 15;44(6):867-873. [doi: 10.1086/511864] [Medline: 17304462]
10. Dellit TH, Owens RC, McGowan JE, Gerding DN, Weinstein RA, Burke JP, Infectious Diseases Society of America, Society for Healthcare Epidemiology of America. Infectious diseases society of America and the society for healthcare epidemiology of America guidelines for developing an institutional program to enhance antimicrobial stewardship. Clin Infect Dis 2007 Jan 15;44(2):159-177. [doi: 10.1086/510393] [Medline: 17173212]
11. Dickstein Y, Geffen Y, Leibovici L, Paul M. Comparison of antibiotic susceptibility patterns of bacterial isolates based on time from hospitalization and culture source: implications for hospital antibiograms. Infect Control Hosp Epidemiol 2016 Feb;37(2):212-214. [doi: 10.1017/ice.2015.258] [Medline: 26564427]
12. Sanden L, Paul M, Leibovici L, Andreassen S. Quantifying the associations between antibiotic exposure and resistance - a step towards personalised antibiograms. Eur J Clin Microbiol Infect Dis 2016 Dec;35(12):1989-1996. [doi: 10.1007/s10096-016-2751-4] [Medline: 27566688]
13. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med 2019 Jul;25(7):1143-1152 [FREE Full text] [doi: 10.1038/s41591-019-0503-6] [Medline: 31273328]
14. Martínez-Agüero S, Mora-Jiménez I, Lérida-García J, Álvarez-Rodríguez J, Soguero-Ruiz C. Machine learning techniques to identify antimicrobial resistance in the intensive care unit. Entropy (Basel) 2019 Jun 18;21(6):603 [FREE Full text] [doi: 10.3390/e21060603] [Medline: 33267317]
15. Zhou H. Large-scale prediction of patient-level antibiotic resistance: towards clinical decision support for improved antimicrobial stewardship. Massachusetts Institute of Technology. 2018. URL: https://dspace.mit.edu/bitstream/handle/1721.1/121646/1098216500-MIT.pdf?sequence=1&isAllowed=y [accessed 2021-08-18]
16. iAntibiogram. University of Pennsylvania Medical Center Guidelines for Antibiotic Use. URL: http://www.uphs.upenn.edu/bugdrug/antibiotic_manual/susceptibility/sustabpmc.htm [accessed 2021-08-18]
17. Curtis C, Al Bahar F, Marriott J. The effectiveness of computerised decision support on antibiotic use in hospitals: a systematic review. PLoS One 2017 Aug 24;12(8):e0183062 [FREE Full text] [doi: 10.1371/journal.pone.0183062] [Medline: 28837665]

18. Rawson TM, Moore LS, Hernandez B, Charani E, Castro-Sanchez E, Herrero P, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? Clin Microbiol Infect 2017 Aug;23(8):524-532 [FREE Full text] [doi: 10.1016/j.cmi.2017.02.028] [Medline: 28268133]
19. Laka M, Milazzo A, Merlin T. Can evidence-based decision support tools transform antibiotic management? A systematic review and meta-analyses. J Antimicrob Chemother 2020 May 01;75(5):1099-1111. [doi: 10.1093/jac/dkz543] [Medline: 31960021]
20. Binkley S, Fishman NO, LaRosa LA, Marr AM, Nachamkin I, Wordell D, et al. Comparison of unit-specific and hospital-wide antibiograms: potential implications for selection of empirical antimicrobial therapy. Infect Control Hosp Epidemiol 2006 Jul;27(7):682-687. [doi: 10.1086/505921] [Medline: 16807842]
21. Jorgensen S, Zurayk M, Yeung S, Terry J, Dunn M, Nieberg P, et al. Emergency department urinary antibiograms differ by specific patient group. J Clin Microbiol 2017 Sep;55(9):2629-2636 [FREE Full text] [doi: 10.1128/JCM.00481-17] [Medline: 28615465]
22. Liu L, Ni Y. Antimicrobial susceptibilities of specific syndromes created with organ-specific weighted incidence antibiograms (OSWIA) in patients with intra-abdominal infections. BMC Infect Dis 2018 Nov 19;18(1):584 [FREE Full text] [doi: 10.1186/s12879-018-3494-x] [Medline: 30453893]
23. Simpao A, Ahumada L, Martinez BL, Cardenas A, Metjian T, Sullivan K, et al. Design and implementation of a visual analytics electronic antibiogram within an electronic health record system at a tertiary pediatric hospital. Appl Clin Inform 2018 Jan;9(1):37-45 [FREE Full text] [doi: 10.1055/s-0037-1615787] [Medline: 29342478]
24. Leibovici L, Fishman M, Schonheyder H, Riekehr C, Kristensen B, Shraga I, et al. A causal probabilistic network for optimal treatment of bacterial infections. IEEE Trans Knowl Data Eng 2000;12(4):517-528. [doi: 10.1109/69.868905]
25. Leibovici L, Kariv G, Paul M. Long-term survival in patients included in a randomized controlled trial of TREAT, a decision support system for antibiotic treatment. J Antimicrob Chemother 2013 Nov;68(11):2664-2666. [doi: 10.1093/jac/dkt222] [Medline: 23743088]
26. Tallman G, Vilches-Tran R, Elman M, Bearden D, Taylor J, Gorman P, et al. Empiric antibiotic prescribing decisions among medical residents: the role of the antibiogram. Infect Control Hosp Epidemiol 2018 May;39(5):578-583. [doi: 10.1017/ice.2018.28] [Medline: 29493481]
27. Norman DA, Stappers PJ. DesignX: complex sociotechnical systems. She Ji: J Des Econ Innov 2015;1(2):83-106. [doi: 10.1016/j.sheji.2016.01.002]
28. Heer J. Agency plus automation: designing artificial intelligence into interactive systems. Proc Natl Acad Sci U S A 2019 Feb 05;116(6):1844-1850 [FREE Full text] [doi: 10.1073/pnas.1807184115] [Medline: 30718389]
29. Cressman A, MacFadden DR, Verma A, Razak F, Daneman N. Empiric antibiotic treatment thresholds for serious bacterial infections: a scenario-based survey study. Clin Infect Dis 2019 Aug 30;69(6):930-937. [doi: 10.1093/cid/ciy1031] [Medline: 30535310]
30. WHO releases the 2019 AWaRe Classification Antibiotics. World Health Organization. URL: https://www.who.int/news/item/01-10-2019-who-releases-the-2019-aware-classification-antibiotics [accessed 2021-08-18]
31. Lovis C, Douglas T, Pasche E, Ruch P, Colaert D, Stroetmann K. DebugIT: building a European distributed clinical data mining network to foster the fight against microbial diseases. Stud Health Technol Inform 2009;148:50-59. [Medline: 19745235]
32. Segura B, Morales AM, Juarez J, Campos M, Palacios F. WASPSS: a clinical decision support system for antimicrobial stewardship. In: Sartipi K, Edoh T, editors. Recent Advances in Digital System Diagnosis and Management of Healthcare. London, UK: Intech Open; Apr 1, 2020:1-148.
33. Hernandez B, Herrero P, Rawson T, Moore L, Charani E, Holmes A, et al. Data-driven web-based intelligent decision support system for infection management at point-of-care: case-based reasoning benefits and limitations. In: Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - HEALTHINF. 2017 Presented at: 10th International Joint Conference on Biomedical Engineering Systems and Technologies - HEALTHINF; 2017; Porto, Portugal p. 119-127. [doi: 10.5220/0006148401190127]
34. Rawson T, Hernandez B, Moore L, Herrero P, Charani E, Ming D, et al. A real-world evaluation of a case-based reasoning algorithm to support antimicrobial prescribing decisions in acute care. Clin Infect Dis 2021 Jun 15;72(12):2103-2111. [doi: 10.1093/cid/ciaa383] [Medline: 32246143]
35. Cetinkaya Y, Falk P, Mayhall CG. Vancomycin-resistant enterococci. Clin Microbiol Rev 2000 Oct;13(4):686-707 [FREE Full text] [doi: 10.1128/CMR.13.4.686] [Medline: 11023964]
36. Los Angeles county regional antibiogram. County of Los Angeles Public Health. URL: http://publichealth.lacounty.gov/acd/AntibiogramData.htm [accessed 2021-08-18]
37. Antibiograms. Washington State Department of Health. URL: https://www.doh.wa.gov/ForPublicHealthandHealthcareProviders/HealthcareProfessionsandFacilities/HealthcareAssociatedInfections/AntibioticResistance/Stewardship/Antibiograms [accessed 2021-08-18]
38. Hostler CJ, Moehring RW, Ashley ES, Johnson M, Davis A, Lewis SS, CDC Prevention Epicenters Program. Feasibility and value of developing a regional antibiogram for community hospitals. Infect Control Hosp Epidemiol 2018 Jun;39(6):718-722 [FREE Full text] [doi: 10.1017/ice.2018.71] [Medline: 29681253]

39. Stelling J, O'Brien TF. Surveillance of antimicrobial resistance: the WHONET program. Clin Infect Dis 1997 Jan;24 Suppl 1:157-168. [doi: 10.1093/clinids/24.supplement_1.s157] [Medline: 8994799]

Abbreviations

AMR: antimicrobial resistance
CDSS: clinical decision support system
CRE: carbapenem-resistant Enterobacteriaceae
CRPA: carbapenem-resistant Pseudomonas aeruginosa
EHR: electronic health record
ICD-10: International Classification of Diseases-10
LOS: length of stay
OR: odds ratio
UCSDHS: University of California San Diego Health Sciences
UTI: urinary tract infection

Edited by R Kukafka, G Eysenbach; submitted 16.08.20; peer-reviewed by P Dutey-Magni, S Sarbadhikari; comments to author 19.09.20; revised version received 30.10.20; accepted 02.08.21; published 03.12.21.

Please cite as:
Müller L, Srinivasan A, Abeles SR, Rajagopal A, Torriani FJ, Aronoff-Spencer E
A Risk-Based Clinical Decision Support System for Patient-Specific Antimicrobial Therapy (iBiogram): Design and Retrospective Analysis
J Med Internet Res 2021;23(12):e23571
URL: https://www.jmir.org/2021/12/e23571
doi:10.2196/23571
PMID:34870601

Review

Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-analysis

Chang Seok Bang[1,2,3,4], MD, PhD; Jae Jun Lee[3,4,5], MD; Gwang Ho Baik[1,2], MD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Republic of Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea

[5]Department of Anesthesiology and Pain Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
77 Sakju-ro
Chuncheon, 24253
Republic of Korea
Phone: 82 33 240 5821
Fax: 82 33 241 8064
Email: csbang@hallym.ac.kr

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/1/e36170

Abstract

Background: Interpretation of capsule endoscopy images or movies is operator-dependent and time-consuming. As a result, computer-aided diagnosis (CAD) has been applied to enhance the efficacy and accuracy of the review process. Two previous meta-analyses reported the diagnostic performance of CAD models for gastrointestinal ulcers or hemorrhage in capsule endoscopy. However, insufficient systematic reviews have been conducted, which cannot determine the real diagnostic validity of CAD models.

Objective: To evaluate the diagnostic test accuracy of CAD models for gastrointestinal ulcers or hemorrhage using wireless capsule endoscopic images.

Methods: We conducted core databases searching for studies based on CAD models for the diagnosis of ulcers or hemorrhage using capsule endoscopy and presenting data on diagnostic performance. Systematic review and diagnostic test accuracy meta-analysis were performed.

Results: Overall, 39 studies were included. The pooled area under the curve, sensitivity, specificity, and diagnostic odds ratio of CAD models for the diagnosis of ulcers (or erosions) were .97 (95% confidence interval, .95–.98), .93 (.89–.95), .92 (.89–.94), and 138 (79–243), respectively. The pooled area under the curve, sensitivity, specificity, and diagnostic odds ratio of CAD models for the diagnosis of hemorrhage (or angioectasia) were .99 (.98–.99), .96 (.94–0.97), .97 (.95–.99), and 888 (343–2303), respectively. Subgroup analyses showed robust results. Meta-regression showed that published year, number of training images, and target disease (ulcers vs erosions, hemorrhage vs angioectasia) was found to be the source of heterogeneity. No publication bias was detected.

Conclusions: CAD models showed high performance for the optical diagnosis of gastrointestinal ulcer and hemorrhage in wireless capsule endoscopy.

(J Med Internet Res 2021;23(12):e33267) doi:10.2196/33267

KEYWORDS

artificial intelligence; computer-aided diagnosis; capsule endoscopy; ulcer; hemorrhage; gastrointestinal; endoscopy; review; accuracy; meta-analysis; diagnostic; performance; machine learning; prediction models

Introduction

Wireless capsule endoscopy (WCE) allows the investigation of gastrointestinal mucosal lesions in a noninvasive manner. This provides approximately 50,000 to 60,000 video frames and allows visualization of the entire gastrointestinal mucosa in a single examination without causing discomfort to patients or risk of procedure-related adverse events [1,2]. Given that the small intestine has been a blind spot for gastroenterologists, WCE has become the standard investigation modality for obscure gastrointestinal hemorrhage and a widely accepted method for the assessment of small intestinal ulcers or tumors [1]. Despite easy accessibility, safety, and patients' comfort for the examination, WCE has a limitation regarding the interpretation. A tedious reading time of approximately 30 to 120 minutes is required, and a small number of abnormal video frames can be easily mistaken for a normal mucosa by endoscopists [1-3].

Artificial intelligence technology has been adopted in gastrointestinal endoscopy, and the automatic detection or diagnosis of abnormal lesions on endoscopic images or movies has been widely investigated [4,5]. The main benefit of the application of artificial intelligence would be the reduction of the laborious reading time and miss rate of important findings in WCE. Another advantage would be the highly accurate diagnostic performance, which is comparable to that of an endoscopist [6]. These artificial intelligence models are expected to aid in the automatic detection of important lesions in WCE images, thus making it possible to perform automatic reading and interpretation of the entire examination.

Previous studies have reported the performance of computer-aided diagnosis (CAD) models using artificial intelligence in WCE [7,8]. Machine learning– or deep learning–based artificial intelligence models with potential benefits have been reported in these studies. Based on these findings, 2 meta-analyses have been conducted for the pooled diagnostic performance of deep-learning models or convolutional neural network models for the diagnosis of gastrointestinal hemorrhage or ulcers using WCE [7,8]. However, the first meta-analysis searched only 1 database, and a substantial number of important articles were omitted. Moreover, an inaccurate crude number of true positives (TP), false positives (FP), false negatives (FN), or true negatives (TN) of CAD models in each study was reported [7]. This inaccurate pooled diagnostic performance can mislead the readers. The second meta-analysis searched multiple databases; however, it also did not include several important papers, and only a single medical librarian searched all the databases [8]. The main pitfall was the simple pooling of the sensitivity or specificity in each study without considering the distribution of abnormal lesions among the total included lesions in each study. Moreover, the diagnostic performance for the gastrointestinal ulcers and hemorrhage was not separated but combined into a single outcome, and quality assessments in each included study were also omitted. The method of exploring the reason for the heterogeneity and the assessment of publication bias also adhered to the interventional meta-analysis methodology in both meta-analyses but did not satisfy the diagnostic test accuracy (DTA) meta-analysis methodology. Given that the method of conducting interventional and DTA meta-analyses is different and that a widely accepted standard methodology exists in conducting the DTA meta-analysis, this can also mislead the readers (Table 1). Therefore, systematic reviews conducted thus far have been inadequate, and the real diagnostic validity of CAD models in WCE has not yet been determined. This study aimed to evaluate the DTA of CAD models for gastrointestinal ulcers or hemorrhage using WCE images through the standard methodology.

Table 1. Comparison of previous meta-analyses with the current study.

Parameters	This study	Soffer et al [7]	Mohan et al [8]
Number of included studies	20 studies on gastrointestinal ulcers and 19 studies on gastrointestinal hemorrhage	5 studies on gastrointestinal ulcers and 5 studies on gastrointestinal hemorrhage	9 studies for the diagnosis of gastrointestinal ulcers or hemorrhage (did not perform separate analysis between ulcers and hemorrhage)
Main outcome	Separate diagnostic performance of CAD[a] models for the gastrointestinal ulcers or hemorrhage using WCE[b]	Separate diagnostic performance of CAD models for the gastrointestinal ulcers or hemorrhage using WCE	Pooled diagnostic performance of CAD models for gastrointestinal ulcers and hemorrhage using WCE (not a meta-analysis with DTA[c]; lack of consideration for the prevalence of ulcers or hemorrhage in each study and thus no calculation of TP[d], FP[e], FN[f], or TN[g] in each study)
Search strategy	Search of MEDLINE through PubMed, Web of Science, and the Cochrane Library (2 independent authors searched the databases)	Search of MEDLINE through PubMed (2 independent authors searched the database)	Search of ClinicalTrials.gov, Ovid EBM[h] Reviews, Ovid, Embase, Ovid MEDLINE, Scopus, and Web of Science (a single medical librarian searched all the databases)
Inaccurate calculation (coding) of TP/FP/FN/TN	N/A[i]	Inaccurate calculation detected in the study's figures	Not a meta-analysis with DTA; lack of consideration for the prevalence of ulcers or hemorrhage in each study and thus no calculation of TP, FP, FN, or TN in each study
Determination of the heterogeneity between studies	Correlation coefficient between the logarithm of the sensitivity and specificity, beta of HSROC[j] model, visual examination of the SROC curve	I^2 statistics (DTA meta-analysis did not determine heterogeneity with I^2 statistic)	I^2 statistic (DTA meta-analysis did not determine heterogeneity with I^2 statistics)
Quality assessment	QUADAS-2[k]	QUADAS-2	Not assessed
Publication bias	Deeks funnel plot asymmetry test	Not assessed	Not assessed

[a]CAD: computer-aided diagnosis.

[b]WCE: wireless capsule endoscopy.

[c]DTA: diagnostic test accuracy.

[d]TP: true positive.

[e]FP: false positive.

[f]FN: false negative.

[g]TN: true negative.

[h]EBM: evidence-based medicine.

[i]N/A: not applicable.

[j]HSROC: hierarchical summary receiver operating characteristic.

[k]QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies second version.

Methods

Adherence to the Checklist for Systematic Reviews and Meta-analyses

This study was conducted in accordance with the statement of the PRISMA (Preferred Reporting Items for a Systematic Review and Meta-analysis) of DTA Studies [9]. The study protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO) database before initiation of the systematic review (#CRD42021253454). Approval from the institutional review board of the Chuncheon Sacred Heart Hospital was waived.

Search Strategy for Relevant Literature

The authors established searching formulae using keywords related to the performance of CAD models in the detection of ulcer or hemorrhage using WCE images. Medical Subject Headings (MeSH) terminology keywords were used for the establishment of searching formulae (Textbox 1).

Textbox 1. Literature searching strategy for the core databases.

1. CAD of gastrointestinal ulcers in WCE

- Database: MEDLINE (through PubMed)

#1. “artificial intelligence”[tiab] OR “AI”[tiab] OR “deep learning”[tiab] OR “machine learning”[tiab] OR “computer”[tiab] OR “neural network”[tiab] OR “CNN”[tiab] OR “automatic”[tiab] OR “automated”[tiab]: 532189

#2. “capsule endoscopy”[tiab] OR “capsule endoscopy”[Mesh]: 5110

#3. “ulcer”[tiab] OR “ulcer”[Mesh] OR “erosion”[tiab]: 138857

#4. #1 AND #2 AND #3: 29

#5. #4 AND English[Lang]: 29

- Database: Web of Science

#1. artificial intelligence OR AI OR deep learning OR machine learning OR computer OR neural network OR CNN OR automatic OR automated: 1236876

#2. capsule endoscopy: 3524

#3. ulcer: 33664

#4. #1 AND #2 AND #3: 49

- Database: Cochrane Library

#1. artificial intelligence: ab, ti, kw; OR AI: ab, ti, kw; OR deep learning: ab, ti, kw; OR machine learning: ab, ti, kw; OR computer: ab, ti, kw; OR neural network: ab, ti, kw; OR CNN: ab, ti, kw; OR automatic: ab, ti, kw; OR automated: ab, ti, kw: 60327

#2. MeSH descriptor: [capsule endoscopy] explode all trees: 131

#3. capsule endoscopy: ab, ti, kw: 724

#4. #2 OR #3: 724

#5. MeSH descriptor: [ulcer] explode all trees: 1413

#6. ulcer: ab, ti, kw; OR erosion: ab, ti, kw: 20844

#7. #5 or #6: 20844

#8. #1 and #4 and #7: 2 trials

2. CAD of Gastrointestinal hemorrhage in WCE

- Database: MEDLINE (through PubMed)

#1. “artificial intelligence”[tiab] OR “AI”[tiab] OR “deep learning”[tiab] OR “machine learning”[tiab] OR “computer”[tiab] OR “neural network”[tiab] OR “CNN”[tiab] OR “automatic”[tiab] OR “automated”[tiab]: 532189

#2. “capsule endoscopy”[tiab] OR “capsule endoscopy”[Mesh]: 5110

#3. “bleeding”[tiab] OR “hemorrhage”[Mesh] OR “angioectasia”[tiab]: 475519

#4. #1 AND #2 AND #3: 82

#5. #4 AND English[Lang]: 79

- Database: Web of Science

#1. artificial intelligence OR AI OR deep learning OR machine learning OR computer OR neural network OR CNN OR automatic OR automated: 1236876

#2. capsule endoscopy: 3524

#3. bleeding OR hemorrhage OR angioectasia: 146789

#4 #1 AND #2 AND #3: 87

- Database: Cochrane Library

#1. artificial intelligence: ab, ti, kw; OR AI: ab, ti, kw; OR deep learning: ab, ti, kw or machine learning: ab, ti, kw; OR computer: ab, ti, kw; OR neural network: ab, ti, kw; OR CNN: ab, ti, kw; OR automatic: ab, ti, kw; OR automated: ab, ti, kw: 60327

#2. MeSH descriptor: [capsule endoscopy] explode all trees: 131

#3. capsule endoscopy: ab, ti, kw: 724

#4. #2 or #3: 724

#5. MeSH descriptor: [hemorrhage] explode all trees: 14887

#6. bleeding: ab, ti, kw; OR angioectasia: ab, ti, kw: 46708

#7. #5 OR #6: 53831

#8. #1 AND #4 AND #7: 8 (trials)

Abbreviations

CAD, computer-aided diagnosis; WCE: wireless capsule endoscopy; tiab: searching code for title and abstract; Mesh: Medical Subject Headings; ab, ti, kw: search code for abstract, title, and keywords; Lang: search code for language; lim: searching code by limiting certain conditions.

Two authors, CSB and JJL, independently performed a core database search of MEDLINE through PubMed, Web of Science, and Cochrane Library using pre-established search formulae from inception to May 2021. Duplicate articles were excluded. The titles and abstracts of all identified articles were reviewed, and irrelevant articles were excluded. Full-text reviews were subsequently performed to determine whether the pre-established inclusion criteria were satisfied in the identified literature. The references of relevant articles were also reviewed to identify any additional studies. Any disagreements of results in the searching process between CSB and JJL were resolved by discussion or consultation with the other author (GHB).

Inclusion Criteria of the Literature

The literature included in this systematic review met the following inclusion criteria: designed to evaluate the diagnostic performance of CAD models for gastrointestinal ulcers or hemorrhage based on WCE images; presentation of the diagnostic performance of CAD models, including sensitivity, specificity, likelihood ratios, predictive values, or accuracy, which enabled the estimation of TP, FP, FN, and TN values of CAD models; and written in English. The exclusion criteria were as follows: narrative review articles, studies with incomplete data, systematic review or meta-analyses, comments, proceedings with only an abstract, or study protocols. A full publication with PDF files of available proceedings was considered to be a full article. Articles meeting at least 1 of the exclusion criteria were excluded from this study.

Assessment of Methodological Quality in the Selected Literature

The methodological quality of the included articles was assessed by CSB and JJL using the second version of Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2). This tool comprises 4 domains, including “patient selection,” “index test,” “reference standard,” and “flow and timing,” with the first 3 domains possessing an “applicability” assessment. CSB and JJL assessed each part as having either a high, low, or unclear risk of bias [10].

Data Extraction in the Selected Literature, Primary Outcomes of This Study, and Additional Analyses

CSB and JJL independently extracted the data from each included article and cross-checked the extracted data. If data were unclear, the corresponding author of the study was contacted by email to obtain insight into the original data set. A descriptive synthesis was made by a systematic review process, and DTA meta-analysis was conducted if the included studies were sufficiently homogenous.

The primary outcomes were the TP, FP, FN, and TN values in each study. For the CAD of gastrointestinal ulcers or hemorrhage using WCE images, the primary outcomes were defined as follows: TP, the number of patients with a positive finding by a CAD model and who had ulcers or hemorrhage as evidenced by WCE images; FP, the number of patients with a positive finding by a CAD model and who did not have ulcers or hemorrhage based on WCE images; FN, the number of patients with a negative finding by a CAD model and who had ulcers or hemorrhage as evidenced by WCE images; and TN, number of patients with a negative finding by a CAD model and who did not have ulcers or hemorrhage based on the WCE images. With these definitions, TP, FP, FN, and TN values were calculated for each included study.

For additional analysis, such as subgroup analysis or meta-regression, the authors extracted the following variables from each included study: published year, geographic origin of the data (ie, Western vs Asian vs public data or unknown), type of CAD models, type of endoscopic images, number of total images included, type of test data sets (internal test vs external test), and the target disease (ulcer vs erosions, hemorrhage, or angioectasias).

Statistics

The hierarchical summary receiver operating characteristic (HSROC) method was used for the DTA meta-analysis [11]. A forest plot of the sensitivity and specificity and a SROC curve were generated. The level of heterogeneity across the included articles was determined by the correlation coefficient between logit-transformed sensitivity and specificity by the bivariate method [12]; for this, the asymmetry parameter was β, where β=0 corresponds to a symmetric ROC curve in which the diagnostic odds ratio (DOR) does not vary along the curve according to the HSROC method [11]. A positive correlation coefficient and a β value with a significant probability (P<.05) indicate heterogeneity between the studies [11,13]. Visual examination of the SROC curve was also performed to identify heterogeneity. Subgroup analysis by univariate meta-regression using the modifiers identified during the systematic review was also performed to identify the reasons for heterogeneity. STATA software version 15.1 (StataCorp), including the packages “metandi” and “midas,” was used for the DTA meta-analysis. Publication bias was evaluated using the Deeks funnel plot asymmetry test. For the subgroup analyses including less than 4 studies, the Moses-Shapiro-Littenberg method [14] was

adopted using Meta-DiSc 1.4 (XI Cochrane Colloquium) because the "metandi" and "midas" packages require the inclusion of a minimum of 4 studies for DTA meta-analysis.

Results

Study Selection

A total of 254 studies (80 studies for the CAD of gastrointestinal ulcers or erosions and 174 studies for the CAD of gastrointestinal hemorrhage using WCE) were identified following a literature search of 3 databases. Fifteen studies were additionally identified by manual screening of bibliographies. After excluding duplicate studies, additional articles were excluded after a review of titles and abstracts. Full-text versions of the remaining 54 and 118 articles were obtained and thoroughly reviewed based on the aforementioned inclusion and exclusion criteria in each topic. Among these, 133 articles were excluded from the final enrollment. Finally, 20 studies [15-34] for the CAD of gastrointestinal ulcers or erosions and 19 [17,19,24,34-49] studies for the diagnosis of gastrointestinal hemorrhage were included in the systematic review. A flowchart of the selection process is shown in Figure 1 and 2.

Figure 1. Flowchart of the search process for the diagnostic performance of computer-aided diagnosis for gastrointestinal ulcers or erosions in wireless capsule endoscopy.

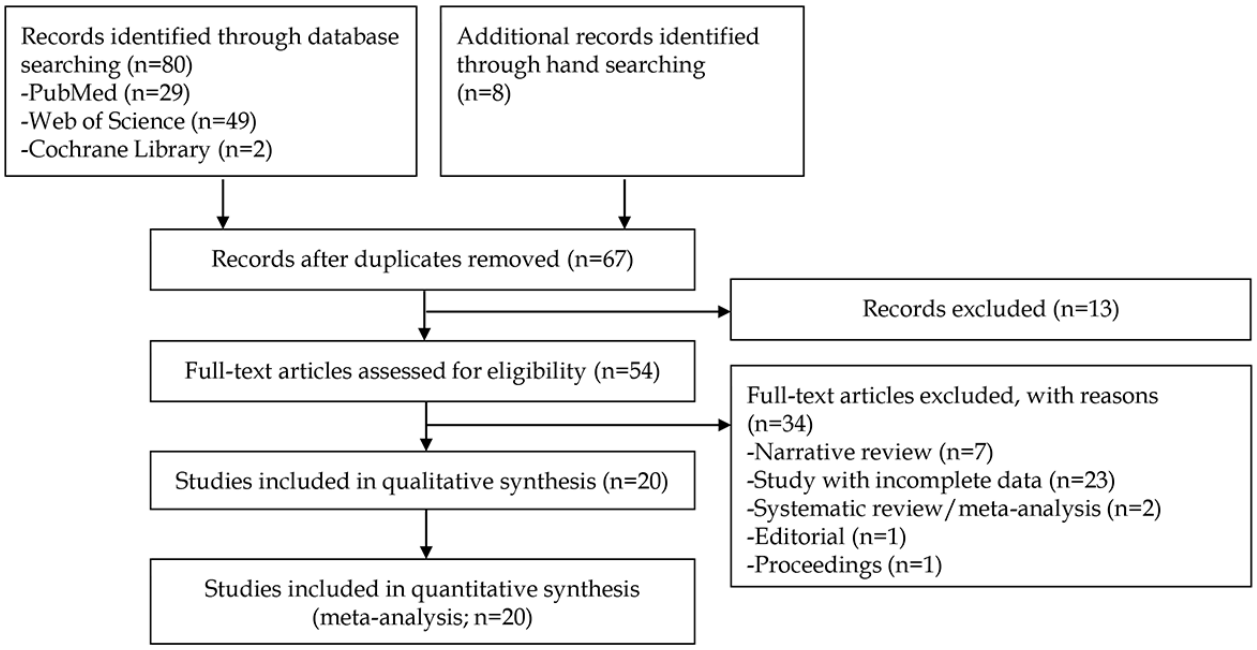
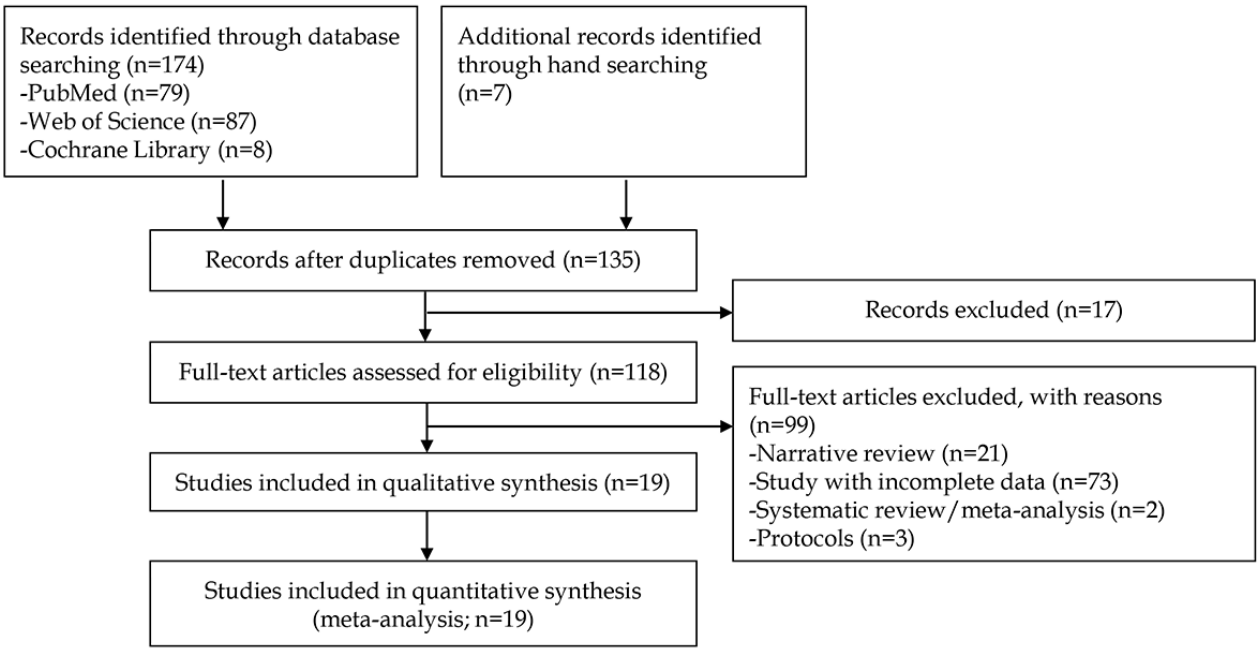


Figure 2. Flowchart of the search process for the diagnostic performance of computer-aided diagnosis for the gastrointestinal hemorrhage in wireless capsule endoscopy.



Clinical Features in the Included Studies

Among the 20 studies [15-34] for the CAD of gastrointestinal ulcers or erosions using WCE, a total of 40,809 images were identified (14,866 cases vs 25,943 controls) for the assessment of the diagnostic performance. Given that the duplicate data were identified (Karargyris et al in 2009 [15] and Karargyris et al in 2011 [20]), all the analyses used the data of 19 studies [15-19,21-34] (a study by Karargyris et al in 2011 [20] was omitted in the meta-analysis).

Ten studies [16-18,21,25-27,29,31,32] used endoscopic images from Asian populations, and 3 studies [22,23,33] used endoscopic images from Western populations. However, 6 studies [15,19,24,28,30,34] used public database images or an unknown source of images. Regarding the type of CAD model, a deep neural network or convolutional neural network was used in 9 studies [16-18,27-29,31-33], and machine learning–based models were used in 10 studies [15,19,21-26,30,34]. Most of the included studies [15-28,30-34] presented the diagnostic performance for the intestinal ulcers. However, the study by Aoki et al [29] presented an indistinguishable performance for the intestinal ulcers or erosions, and the study by Fan et al [27] presented a separate performance for the intestinal ulcers and erosions. Therefore, subgroup analyses were performed for the target lesions. Detailed clinical features of the included studies are presented in Multimedia Appendix 1.

Among the 19 studies [17,19,24,34-49] for the diagnosis of gastrointestinal hemorrhage using WCE, a total of 41,323 images were identified (6952 cases vs 34,371 controls) for the assessment of the diagnostic performance.

Five studies [17], [35], [44], [48], [49] used endoscopic images from Asian populations, and 1 study [47] used endoscopic images from Western populations. However, the remaining 13 studies [19,24,34,36-43,45,46] used public database images or an unknown source of images. Regarding the type of CAD model, the deep neural network or convolutional neural network was used in 8 studies [17,35,39,43,46-49], and machine learning–based models were used in 11 studies [19,24,34,36-38,40-42,44,45]. Most of the included studies [17,19,24,34-46,48] presented the diagnostic performance for intestinal hemorrhage. However, studies by Leenhardt et al [47] and Tsuboi et al [49] presented the performance for angiodysplasias. Therefore, subgroup analyses were performed for the target lesions. Detailed clinical features of the included studies are presented in Multimedia Appendix 2.

Quality Assessment of Study Methodology

The quality and quantity of baseline training data are important because the CAD models are established using learning features of the baseline training data. A sufficient number of training images are required for the establishment of practical CAD models, and endoscopic specialists should participate in the labeling work for the accurate preparation of training data. We also could not guarantee the quality of images in public databases searched on the internet. We determined that proper learning requires at least 30 training images (quantity standard) from real clinical hospital data (quality standard) labeled by an endoscopic specialist (quality standard). If both quality and quantity standards were satisfied, there was considered to be a low risk of bias in the patient selection domain. If only 1 of these quality or quantity standards was satisfied, there was considered to be an unclear risk of bias. If both were not satisfied, there was considered to be a high risk of bias.

In terms of the CAD of gastrointestinal ulcers or erosions in WCE, only 7 studies [25-27,29,31-34] were rated as low risk of bias, 9 studies [16-18,21-24,30,34] were rated as unclear risk of bias, and 3 studies [15,19,28] were rated as high risk of bias in the “patient selection” domain. The remaining domains were rated as having a low risk of bias in all the included studies (Figure 3). Therefore, classification of methodological quality in the “patient selection” domain was adopted as a modifier in the subgroup or meta-regression analysis.

In terms of the CAD of gastrointestinal hemorrhage in WCE, only 3 studies [44,47,49] were rated as having a low risk of bias, 10 studies [17,24,34,35,37,38,40,45,46,48] were rated as having an unclear risk of bias, and 6 studies [19,36,39,41-43] were rated as having a high risk of bias in the “patient selection” domain. The remaining domains were rated as having a low risk of bias in all the included studies (Figure 4). Therefore, the classification of methodological quality in the “patient selection” domain was adopted as a modifier in the subgroup or meta-regression analysis.

Figure 3. Summary graph of quality in methodology for the computer-aided diagnosis of gastrointestinal ulcers or erosions in wireless capsule endoscopy.

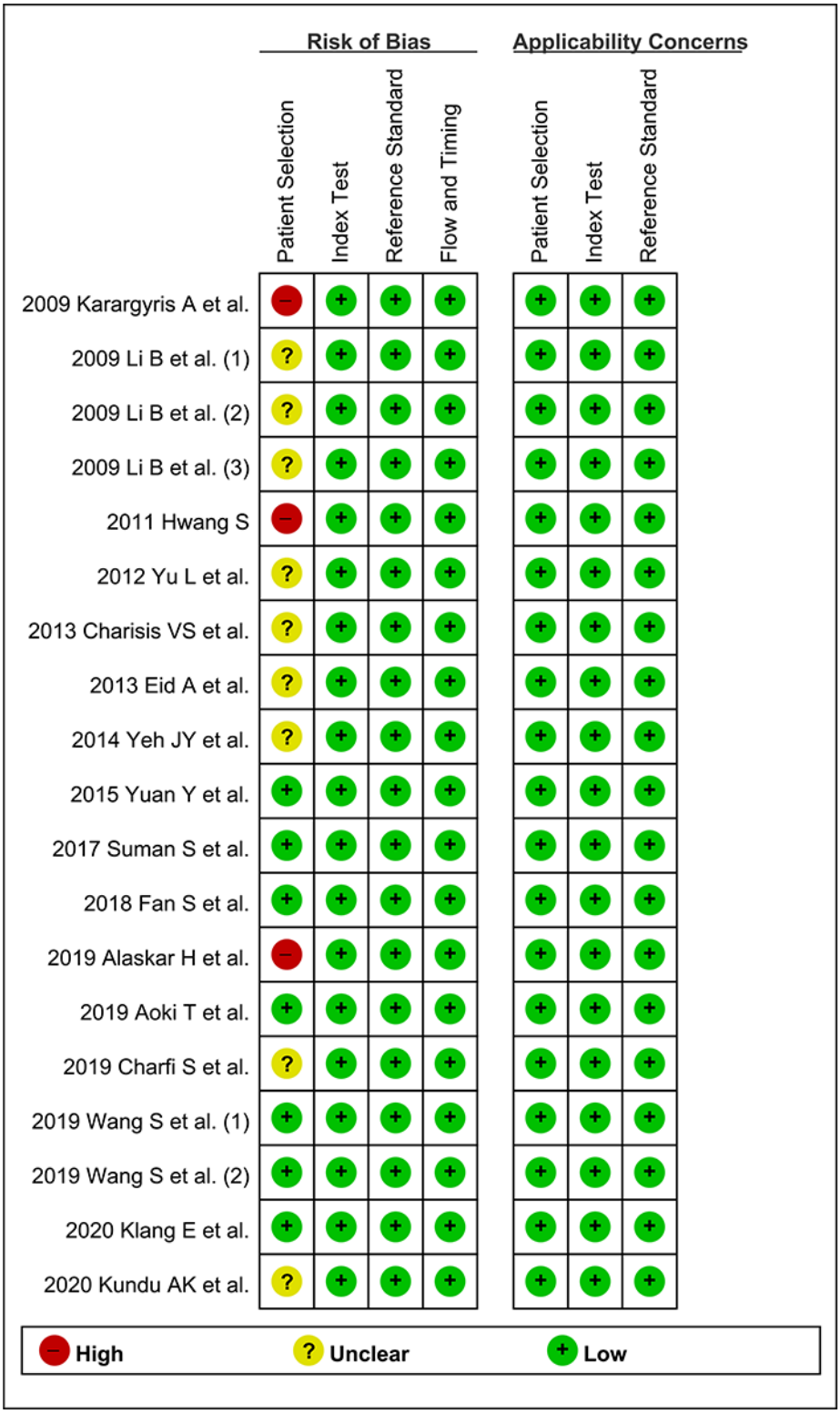
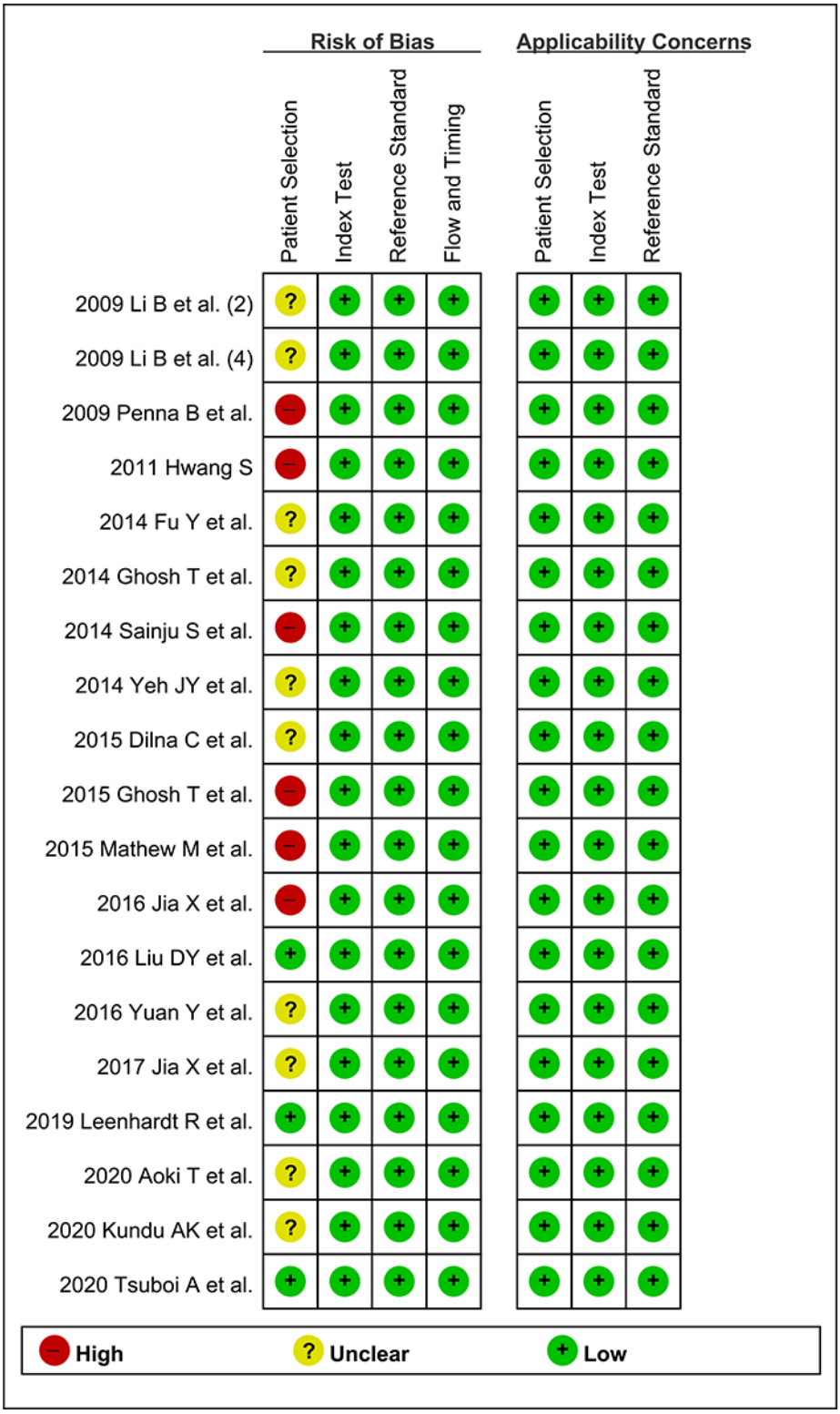


Figure 4. Summary graph of quality in methodology for the computer-aided diagnosis of gastrointestinal hemorrhage in wireless capsule endoscopy.



DTA Meta-analysis for the Performance of CAD Models

Among the 20 studies [15-34] for the meta-analysis of the CAD of gastrointestinal ulcers or erosions using WCE, the area under the curve (AUC), sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and DOR were 0.97 (95% CI 0.95-0.98), 0.93 (95% CI 0.89-0.95), 0.92 (95% CI 0.89-0.94), 11.2 (95% CI 8.6-14.7), 0.08 (95% CI 0.05-0.12), and 138 (95% CI 79-243), respectively (Multimedia Appendix 3 and Figure 5). The SROC curve is illustrated in Figure 6. To investigate the clinical utility of the CAD models, Fagan's nomogram [50] was generated. Positive findings indicated that gastrointestinal ulcers or erosions were detected by the CAD models, while negative findings indicated that gastrointestinal ulcers or erosions were not detected by the CAD models. Assuming a 23% prevalence of gastrointestinal ulcers or erosions, Fagan's nomogram showed that the posterior probability of ulcers or

erosions was 76% if the finding of the CAD model was positive and that the posterior probability of ulcers or erosions was only 3% if the finding of the CAD model was negative (Figure 7).

Among the 19 studies [17,19,24,34-49] for the meta-analysis of the CAD of gastrointestinal hemorrhage in WCE, the AUC, sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and DOR were 0.99 (95% CI 0.98-0.99), 0.96 (95% CI 0.94-0.97), 0.97 (95% CI 0.95-0.99), 38.3 (95% CI 19.6-74.8), 0.04 (95% CI 0.03-0.07), and 888 (95% CI 343-2303), respectively (Multimedia Appendix 4 and Figure 8). The SROC curve is illustrated in Figure 9. Positive findings of Fagan's nomogram indicated that gastrointestinal hemorrhage was detected by the CAD models. Negative findings indicated that gastrointestinal hemorrhage was not detected by the CAD models. Assuming a 10% prevalence of small intestinal hemorrhage in all gastrointestinal bleeding [51], Fagan's nomogram showed that the posterior probability of small intestinal hemorrhage was 81% if the finding of the CAD model was positive and that the posterior probability of small intestinal hemorrhage was only 0.5% if the finding of the CAD model was negative (Figure 10).

Figure 5. Coupled forest plots of sensitivity and specificity in computer-aided diagnosis models for the diagnosis of gastrointestinal ulcers or erosions in wireless capsule endoscopy images.

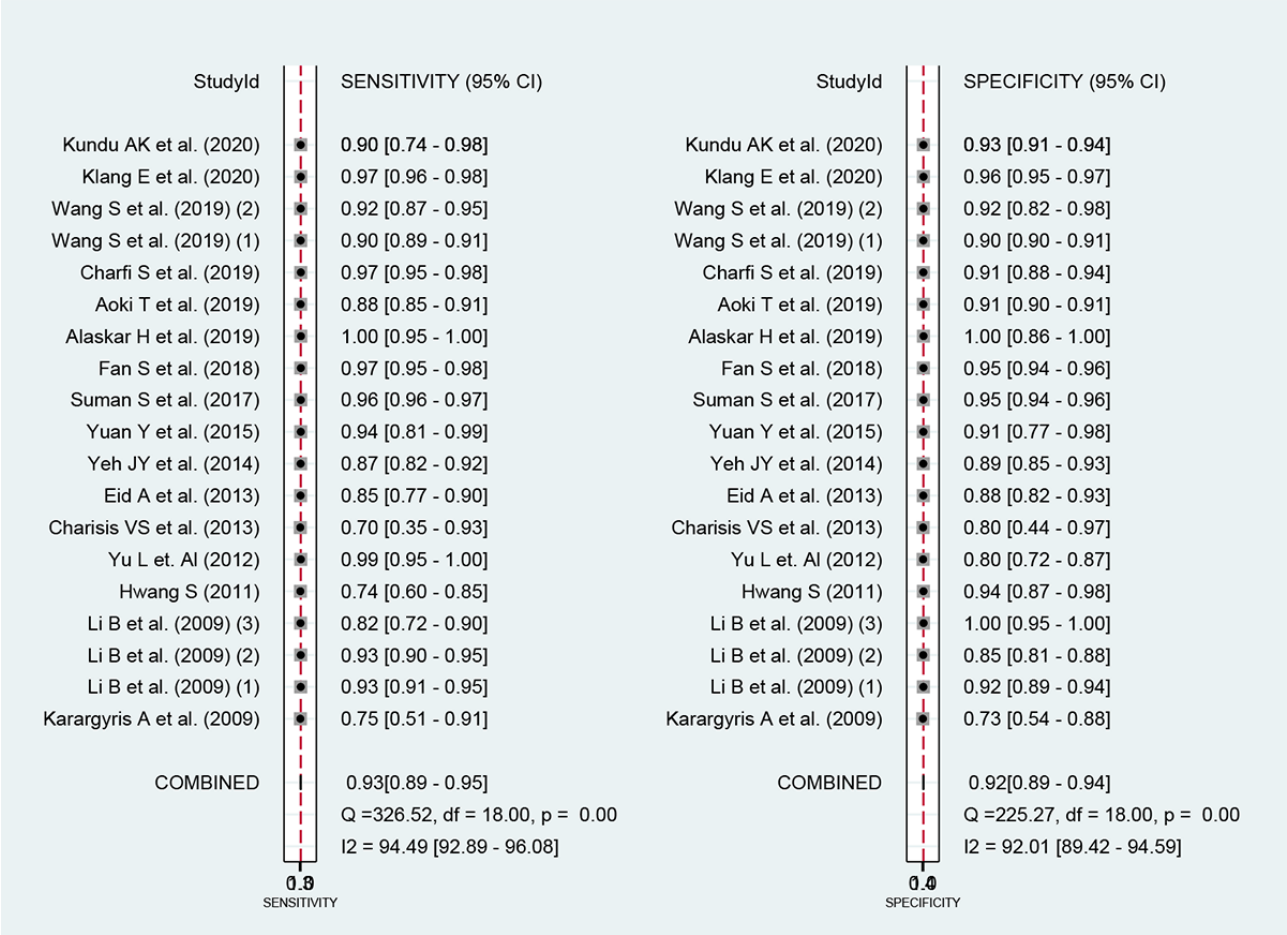


Figure 6. Coupled forest plots of sensitivity and specificity in computer-aided diagnosis models for the diagnosis of gastrointestinal ulcers or erosions in wireless capsule endoscopy images. AUC: area under the curve; SENS: sensitivity; SPEC: specificity; SROC: summary receiver operating characteristic.

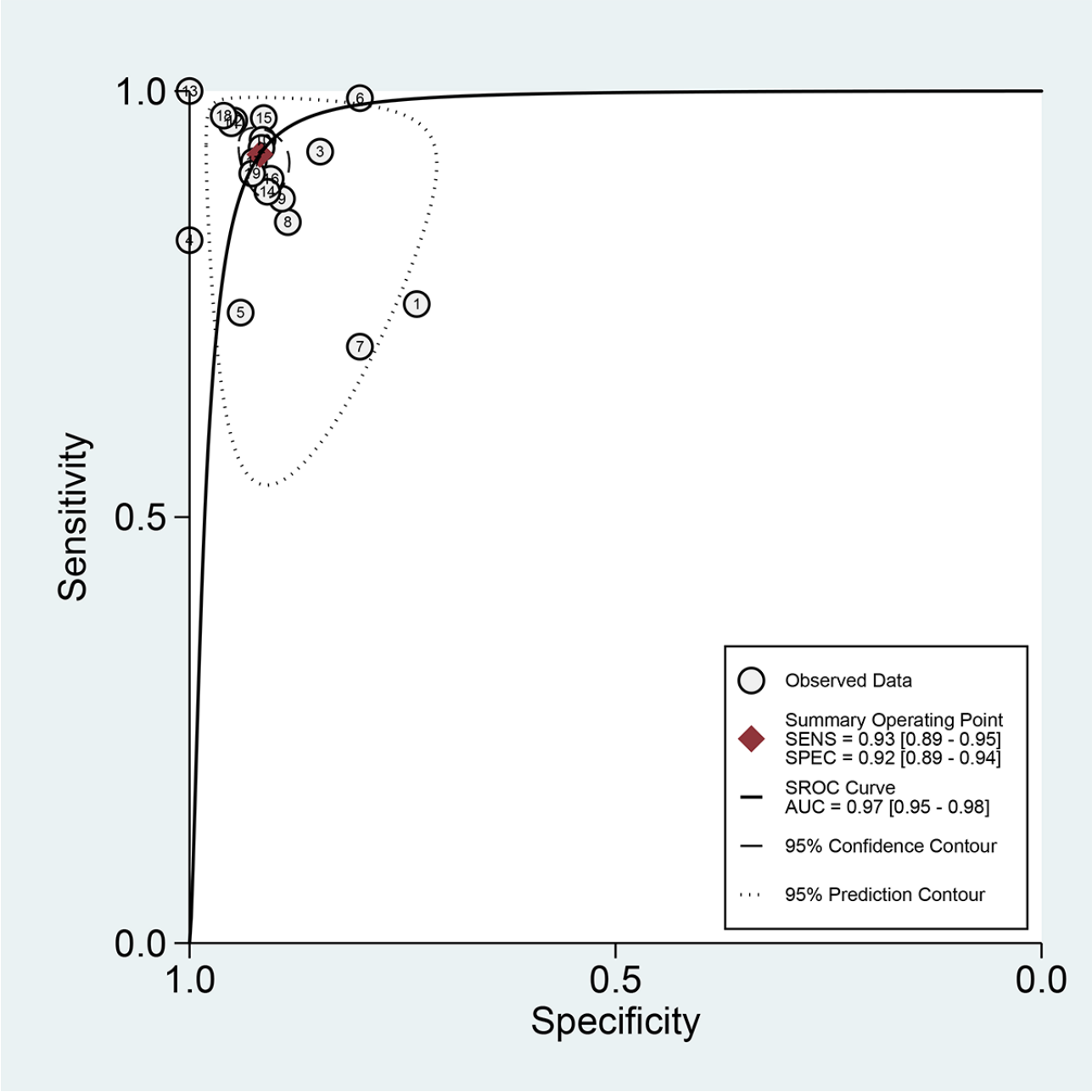


Figure 7. Fagan's nomogram for the computer-aided diagnosis of gastrointestinal ulcers or erosions in wireless capsule endoscopy images.

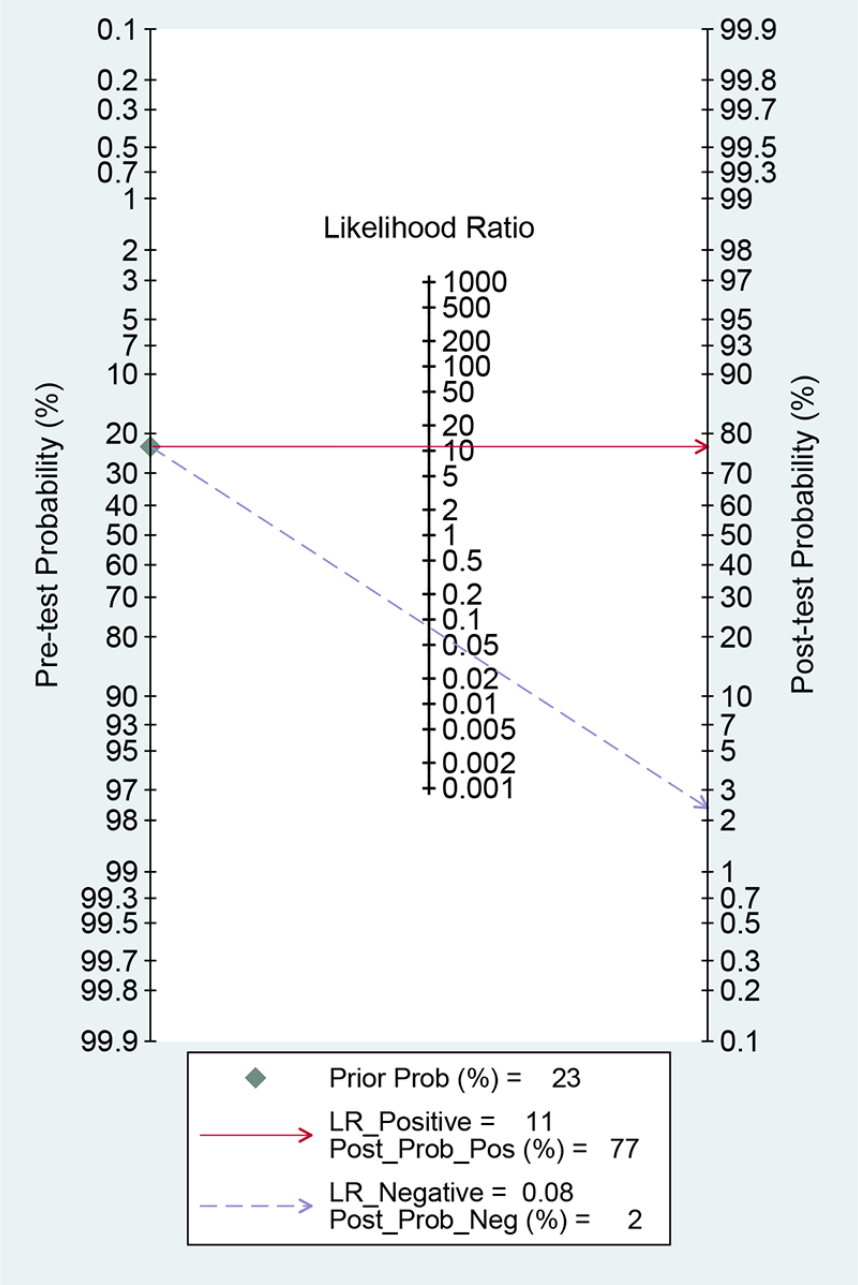


Figure 8. Coupled forest plots of sensitivity and specificity in computer-aided diagnosis models for the diagnosis of gastrointestinal hemorrhage in wireless capsule endoscopy images.

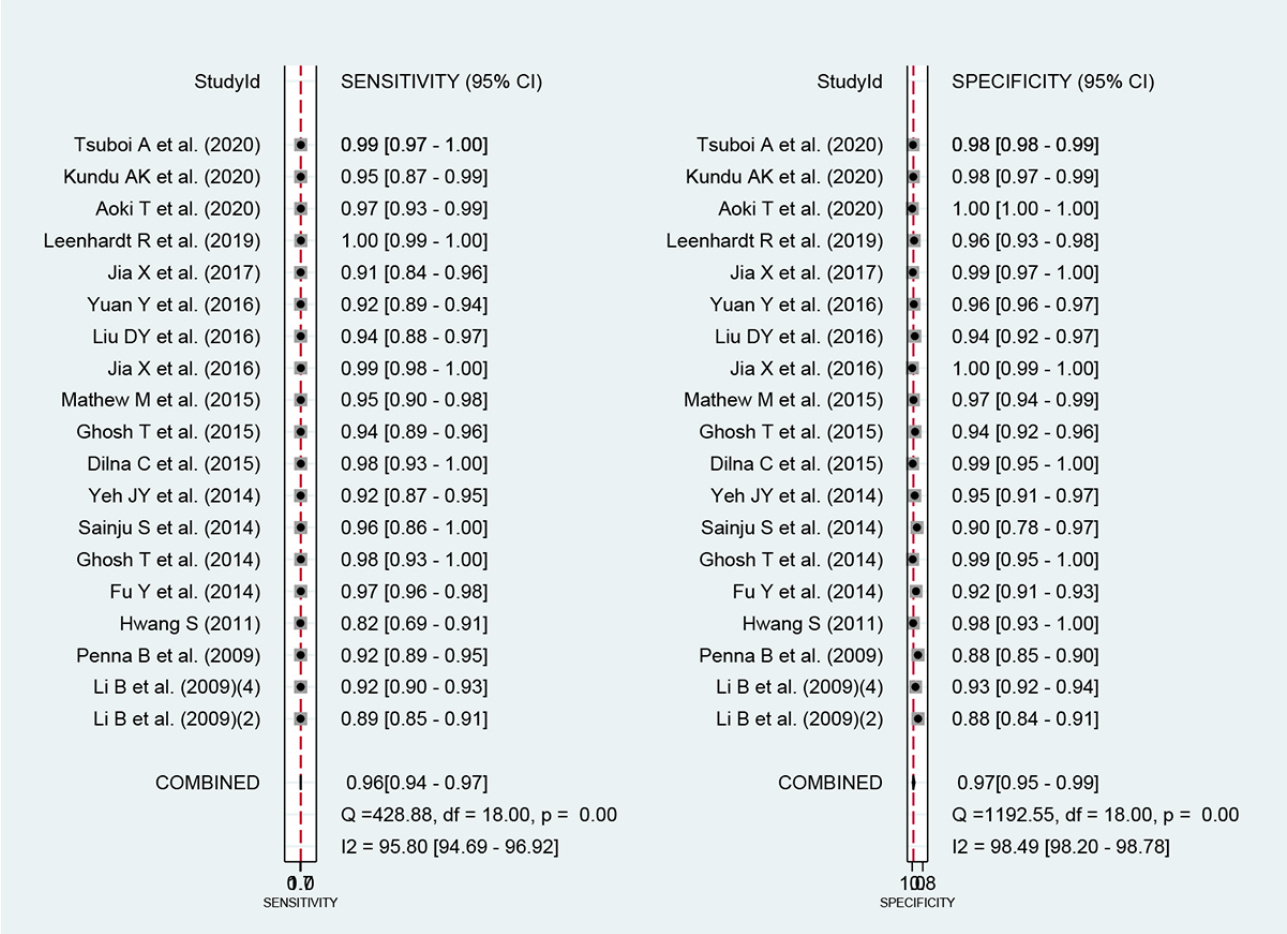


Figure 9. Summary receiver operating characteristic curve with 95% confidence region and prediction region of computer-aided diagnosis models for the diagnosis of gastrointestinal hemorrhage in wireless capsule endoscopy images. AUC: area under the curve; SENS: sensitivity; SPEC: specificity; SROC: summary receiver operating characteristic.

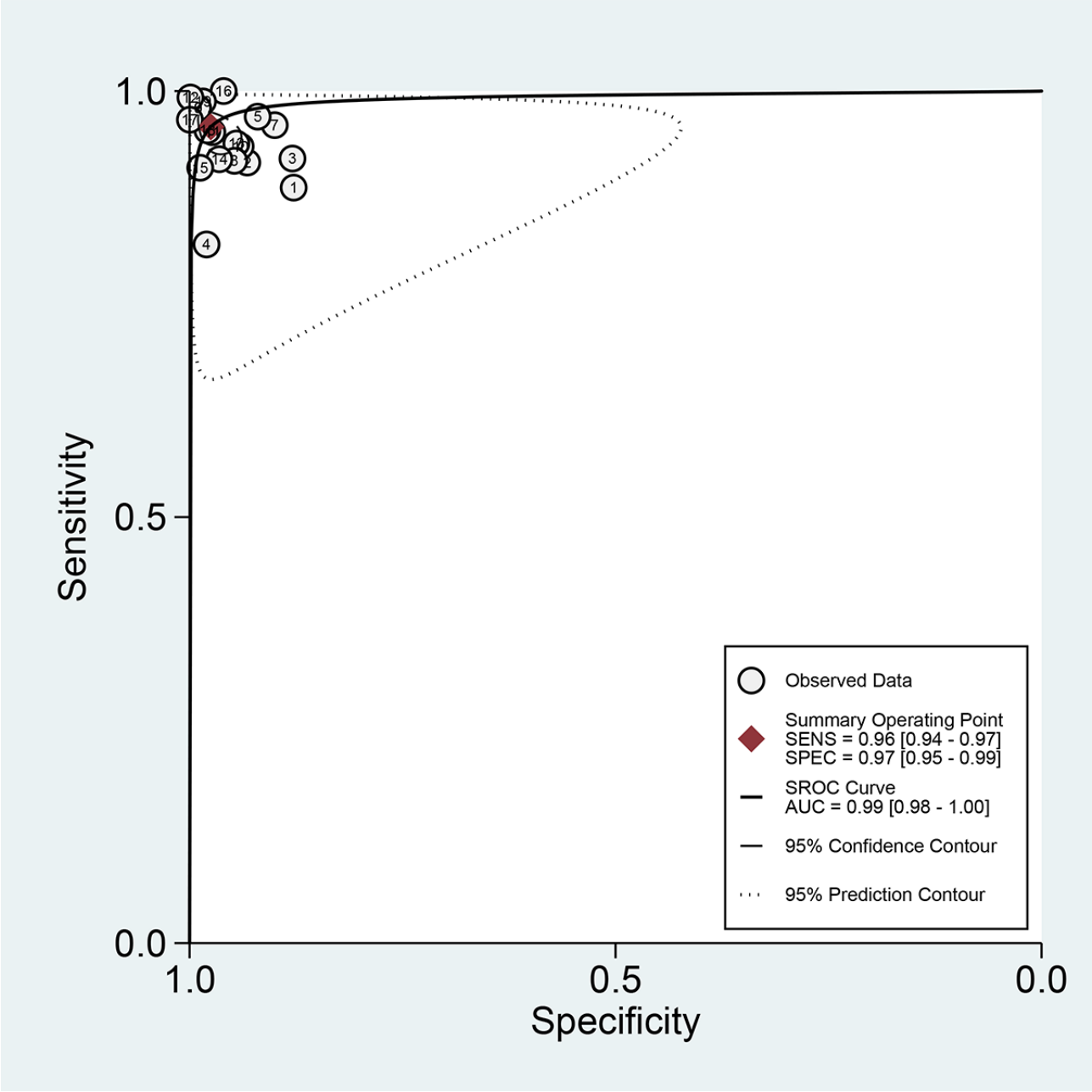
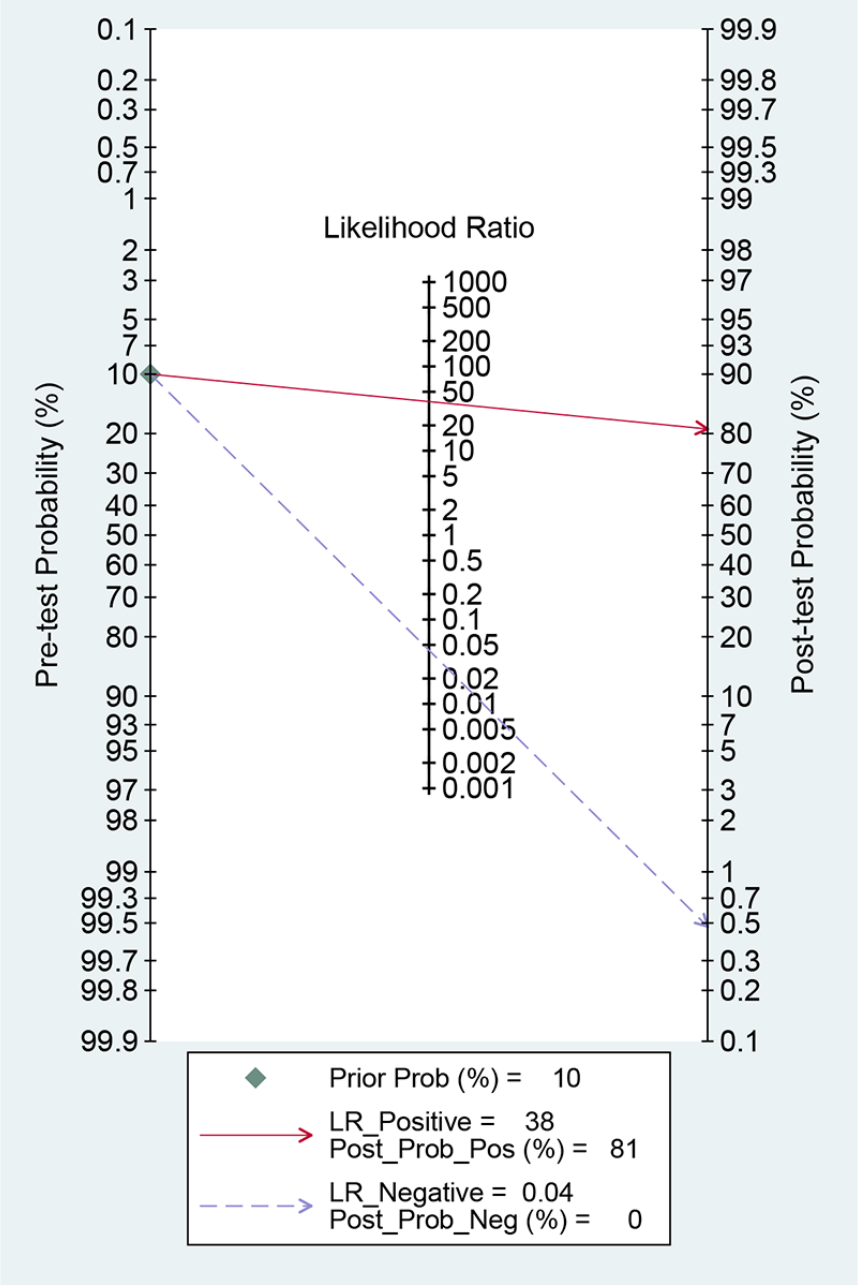


Figure 10. Fagan's nomogram for the computer-aided diagnosis of small intestinal hemorrhage in wireless capsule endoscopy images.



Assessment of Heterogeneity With Meta-regression and Subgroup Analysis

For the CAD of gastrointestinal ulcers or erosions in WCE, we first observed a positive correlation coefficient between the logit-transformed sensitivity and specificity (r=0.28) in the bivariate model analysis. However, an asymmetric β parameter in the HSROC model showed an nonsignificant P value (P=.15), implying that heterogeneity was not present among the studies. Second, a coupled forest plot of sensitivity and specificity was observed (Figure 5). Compared with the enrolled studies, the study by Karargyris et al (2009) [15] showed lower sensitivity and specificity. This study was found to have a high risk of bias in the methodology quality assessment (Figure 3). Therefore, subgroup analysis was carried out according to the methodological quality, and the performance was robust although slightly higher values were observed in the studies of high methodological quality (Multimedia Appendix 3). Third, the shape of the SROC curve for the gastrointestinal ulcers or erosions in WCE was symmetric, and the 95% prediction region was not wide (Figure 6). Fourth, meta-regression using modifiers identified in the systematic review was conducted, and published year, number of training images, and target disease (ulcer vs erosion) were found to be the source of heterogeneity (published year: P=.04; number of training images: P=.02; target disease ulcer vs erosion: P=.38; type of endoscopic image: P=.01). Finally, a subgroup analysis based on the potential modifiers was performed, and the overall performance of the studies published within 10 years (vs studies published more than 10 years ago) and studies with more than 100 training images (vs studies with fewer than 100 training images) showed higher values (Multimedia Appendix 3).

For the CAD of gastrointestinal hemorrhage in WCE, we first observed a positive correlation coefficient between the

logit-transformed sensitivity and specificity (r=0.48) in the bivariate model analysis. However, an asymmetric β parameter in the HSROC model showed an nonsignificant P value (P=.06), implying that heterogeneity was not present among the studies. Second, a coupled forest plot of sensitivity and specificity was observed (Figure 8), and there was no significant outlier. Third, the shape of the SROC curve for the gastrointestinal ulcers and erosions in WCE was symmetric, and the 95% prediction region was not wide (Figure 9). Fourth, a meta-regression using the modifiers identified in the systematic review was conducted, and published year, number of training images, and target disease (hemorrhage vs angioectasia) were found to be the source of heterogeneity (published year: P<.01; number of training images: P=.04; target disease hemorrhage vs angioectasia: P<.01). Finally, a subgroup analysis based on the potential modifiers was performed, and the overall performance of the studies published within 10 years (vs studies published more than 10 years ago) and studies with more than 100 training images (vs studies with fewer than 100 training images) showed higher values (Multimedia Appendix 4).

Evaluation of Publication bias

The Deeks funnel plot of studies for the gastrointestinal ulcers or erosions in WCE exhibited a symmetrical shape with respect to the regression line (Figure 11), and the asymmetry test showed no evidence of publication bias (P=.77). The Deeks funnel plot of studies for the gastrointestinal hemorrhage in WCE exhibited a symmetrical shape with respect to the regression line (Figure 12), and the asymmetry test showed no evidence of publication bias (P=.93).

Figure 11. Deeks funnel plot of computer aided diagnosis models for the diagnosis of gastrointestinal ulcers or erosions in wireless capsule endoscopy images.

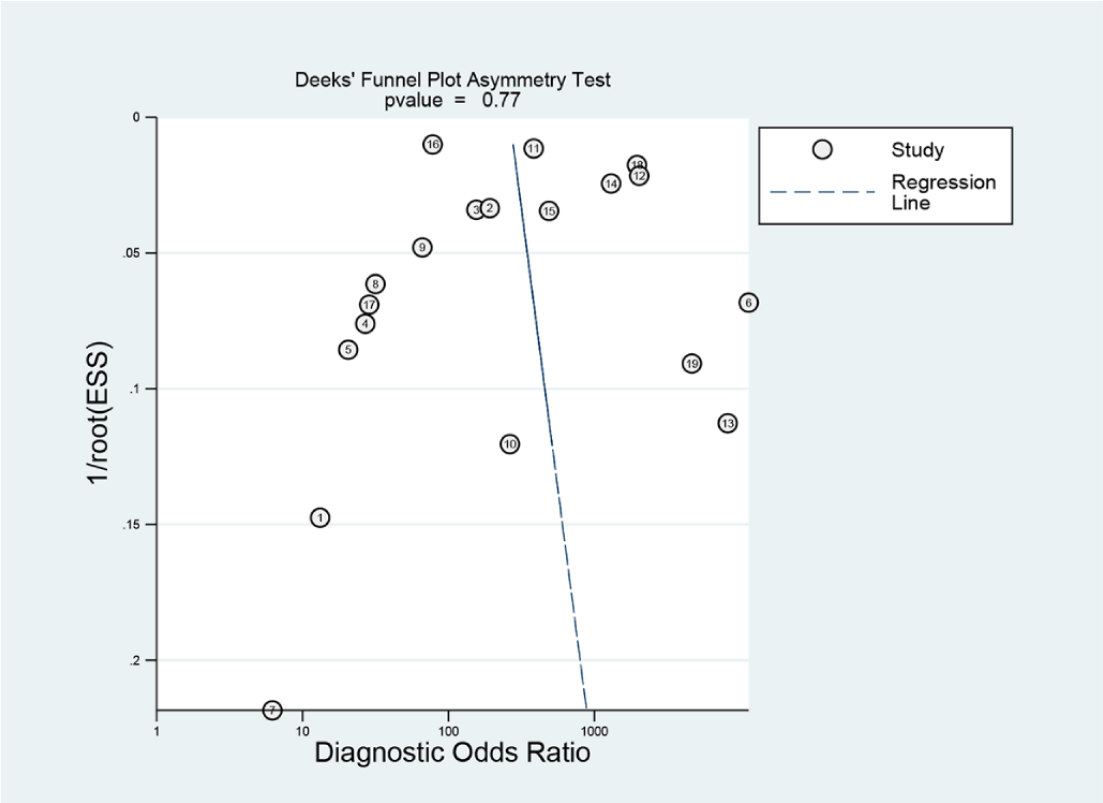
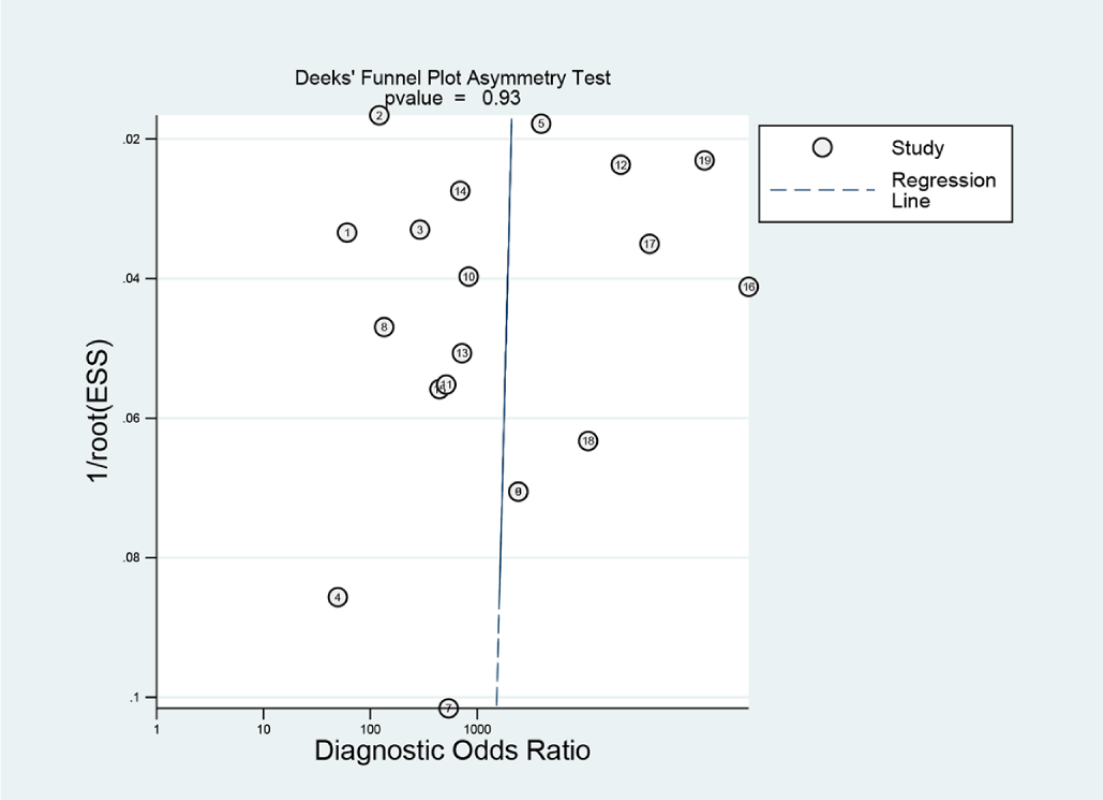


Figure 12. Deeks funnel plot of computer-aided diagnosis models for the diagnosis of gastrointestinal hemorrhage in wireless capsule endoscopy images.



Discussion

Principal Findings

In this study, CAD models showed high performance values for the diagnosis of gastrointestinal ulcers or erosions and hemorrhage in WCE images. Practical values in Fagan's nomogram indicated the potential to use CAD models in clinical practice. Although the main analyses found some heterogeneity among the included studies, the meta-regression showed the common reasons for heterogeneity (published year, number of training images, and target disease—ulcers vs erosions and hemorrhage vs angioectasia), and subgroup analyses demonstrated that recently published studies (vs studies published more than 10 years ago) with a greater amount of training data (vs studies with fewer than 100 training images) showed better performance of CAD models. Thorough subgroup analyses indicated the robust quality of the evidence.

Interpretation of WCE images is an important task for gastroenterologists. Due to the fact that WCE presents the images of the whole gastrointestinal tract, lesions that are difficult to detect with conventional endoscopy can be identified. Diminutive but important culprit lesions also can be found in the WCE examination. The noninvasive nature of this examination and patients' comfort have also promoted the use of this technique in the diagnosis of obscure gastrointestinal hemorrhage or small intestinal disorders. However, the interpretation process is tedious. At least 30 to 120 minutes of reading time is required for the endoscopists [1-3]. It is necessary to maintain concentration throughout the reading time so as not to miss important lesions. CAD models have the potential to automate the reading process of WCE with their high diagnostic performance, especially for sensitivity and specificity. The overall performance is slightly higher for gastrointestinal hemorrhage than for ulcers or erosions. It is presumed that this is because red-colored blood is easier to distinguish than are white- or yellow-colored ulcers or erosions, which are similar to the color of the background mucosa for the pixel-based or red-green-blue spectrum–based feature learning of CAD.

In the context of the learning way of CADs, neural network–based CAD models showed a slightly higher performance than that of traditional machine learning–based CAD models (Multimedia Appendices 3 and 4). CNN is not always better than machine learning for accurate classification. However, image recognition with local feature extraction can be highly optimized with its complex layers and deep nodes calculations and dimensional reductions for neural network CAD models. Considering that the machine learning–based models in the included studies used color or textures features in the images of WCE, neural network–based models might focus on the other local features or combined features, such as the shape of the lesions or feature differences between the lesions and background mucosa. Explainable artificial intelligence analyses are on the rise, and the application of this technique would provide a method of determination in the CAD models [52].

Although meta-analyses of same topic have already been published, this study was conducted to evaluate the DTA of CAD models for gastrointestinal ulcers or hemorrhage using WCE images with a standard methodology (Table 1) [7,8].

Although previous studies also reported the high performance of CAD models, many important articles were omitted, the heterogeneity between studies was determined by I^2 statistic (which is used in interventional meta-analysis), methodological quality assessments were omitted, and the publication bias was also not assessed.

Limitations

Despite the robust evidence in this meta-analysis, several inevitable limitations were identified. First, all the performance data were only measured in an internal-test setting in each included study. Modeling is an assumption that observations follow certain statistical rules, and external validation is a method to check whether this assumption is correct or generalizable. Therefore, the confirmation of performance in the established CAD models with unused data in the training or internal testing process is essential [53]. However, no single study conducted performance verification in an external validation setting. Second, the definition of intestinal ulcers or erosions was vague. Erosion usually refers to damage that is limited to the mucosa (loss of the epithelium but with the basement membrane or lamina propria being intact). However, the definition of ulcers usually involves more extensive loss of the mucosa beyond the lamina propria. Although the discrimination between these 2 conditions is not perfect under visual inspection, there was no clear definition in the included studies. This can lead to the underestimation or overestimation of the performance of CAD models. Third, many studies used baseline training data from a public database, and we could not guarantee the quality of images in the public databases available from the internet. The diagnostic performance of the CAD models can only be valid for the population under evaluation and depends on the prevalence of target conditions for the selected population (so-called spectrum bias or class imbalance) [2,54]. This class imbalance was not considered in the included studies. Most of the studies except for 1 [34] applied a 1:1 to 1:4 ratio (target condition:normal mucosa) of the training data set. However, Kundu et al [34] used 31 ulcer images and 1617 normal mucosal images (about a 1:52 ratio) and 65 bleeding images and 1617 normal mucosal images (about a 1:25 ratio) in the training data set. Considering that the method of establishing artificial intelligence models is changing from a model-centric (ie, change or optimize the model to improve performance) to a data-centric approach (ie, systematically change the distribution of the quality of data to improve performance), model establishment that takes into account spectrum bias is required. Overall, qualified training data with clear definitions and a focus on external validation-oriented performance CAD model establishment are required and expected for future perspectives in this topic.

In conclusion, CAD models showed high performance for the optical diagnosis of gastrointestinal ulcers and hemorrhage in WCE.

Acknowledgments

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (grant number 2020-0-01604).

Authors' Contributions

CSB was responsible for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, writing the original draft, and reviewing and editing the final draft. JJL was responsible for data curation, formal analysis, investigation, and resources. GHB was responsible for data curation, formal analysis, investigation, and resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical characteristics of the included studies for the diagnosis of ulcers or erosions in wireless capsule endoscopy images using computer-aided diagnosis.
[DOCX File , 23 KB - jmir_v23i12e33267_app1.docx]

Multimedia Appendix 2
Clinical characteristics of the included studies for the diagnosis of gastrointestinal hemorrhage in wireless capsule endoscopy images using computer-aided diagnosis.
[DOCX File , 22 KB - jmir_v23i12e33267_app2.docx]

Multimedia Appendix 3
Summary of performance and subgroup analysis of the included studies for the diagnosis of ulcers or erosions in wireless capsule endoscopy images using computer-aided diagnosis.
[DOCX File , 21 KB - jmir_v23i12e33267_app3.docx]

Multimedia Appendix 4
Summary of performance and subgroup analysis of the included studies for the diagnosis of bleeding in wireless capsule endoscopy images using computer-aided diagnosis.
[DOCX File , 18 KB - jmir_v23i12e33267_app4.docx]

References

1. ASGE Technology Committee, Wang A, Banerjee S, Barth BA, Bhat YM, Chauhan S, et al. Wireless capsule endoscopy. Gastrointest Endosc 2013 Dec;78(6):805-815. [doi: 10.1016/j.gie.2013.06.026] [Medline: 24119509]
2. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019 Apr 14;25(14):1666-1683 [FREE Full text] [doi: 10.3748/wjg.v25.i14.1666] [Medline: 31011253]
3. McAlindon ME, Ching H, Yung D, Sidhu R, Koulaouzidis A. Capsule endoscopy of the small bowel. Ann Transl Med 2016 Oct;4(19):369 [FREE Full text] [doi: 10.21037/atm.2016.09.18] [Medline: 27826572]
4. Bang CS. [Deep learning in upper gastrointestinal disorders: status and future perspectives]. Korean J Gastroenterol 2020 Mar 25;75(3):120-131 [FREE Full text] [doi: 10.4166/kjg.2020.75.3.120] [Medline: 32209800]
5. Bang CS, Lim H, Jeong HM, Hwang SH. Use of endoscopic images in the prediction of submucosal invasion of gastric neoplasms: automated deep learning model development and usability study. J Med Internet Res 2021 Apr 15;23(4):e25167 [FREE Full text] [doi: 10.2196/25167] [Medline: 33856356]
6. Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of esophageal cancer and neoplasms in endoscopic images: a systematic review and meta-analysis of diagnostic test accuracy. Gastrointest Endosc 2021 May;93(5):1006-1015.e13 [FREE Full text] [doi: 10.1016/j.gie.2020.11.025] [Medline: 33290771]
7. Soffer S, Klang E, Shimon O, Nachmias N, Eliakim R, Ben-Horin S, et al. Deep learning for wireless capsule endoscopy: a systematic review and meta-analysis. Gastrointest Endosc 2020 Oct;92(4):831-839.e8. [doi: 10.1016/j.gie.2020.04.039] [Medline: 32334015]
8. Mohan BP, Khan SR, Kassab LL, Ponnada S, Chandan S, Ali T, et al. High pooled performance of convolutional neural networks in computer-aided diagnosis of GI ulcers and/or hemorrhage on wireless capsule endoscopy images: a systematic review and meta-analysis. Gastrointest Endosc 2021 Feb;93(2):356-364.e4. [doi: 10.1016/j.gie.2020.07.038] [Medline: 32721487]
9. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163] [Medline: 29362800]
10. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
11. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001 Oct 15;20(19):2865-2884. [doi: 10.1002/sim.942] [Medline: 11568945]
12. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005 Oct;58(10):982-990. [doi: 10.1016/j.jclinepi.2005.02.022] [Medline: 16168343]
13. Harbord RM, Whiting P. Metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. The Stata Journal 2009 Aug 01;9(2):211-229. [doi: 10.1177/1536867x0900900203]
14. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13(4):313-321. [doi: 10.1177/0272989X9301300408] [Medline: 8246704]
15. Karargyris A, Bourbakis N. Identification of ulcers in wireless capsule endoscopy videos. 2009 Presented at: International Symposium on Biomedical Imaging: From Nano to Macro; June 28-July 1, 2009; Bostan, MA, USA. [doi: 10.1109/isbi.2009.5193107]
16. Li B, Meng MQ. Texture analysis for ulcer detection in capsule endoscopy images. Image and Vision Computing 2009 Aug;27(9):1336-1342. [doi: 10.1016/j.imavis.2008.12.003]
17. Li B, Meng MQ. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Comput Biol Med 2009 Mar;39(2):141-147. [doi: 10.1016/j.compbiomed.2008.11.007] [Medline: 19147126]
18. Li B, Qi L, Meng M, Fan Y. Using ensemble classifier for small bowel ulcer detection in wireless capsule endoscopy images. 2009 Presented at: IEEE International Conference on Robotics and Biomimetics; Dec 19-13, 2009; Guilin, Guangxi, China. [doi: 10.1109/robio.2009.5420455]
19. Hwang S. Bag-of-visual-words approach to abnormal image detection in wireless capsule endoscopy videos. 2011 Presented at: International Symposium on Visual Computing: Advances in Visual Computing; Sep 26-28, 2011; Las Vegas, NV, USA. [doi: 10.1007/978-3-642-24031-7_32]
20. Karargyris A, Bourbakis N. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. IEEE Trans. Biomed. Eng 2011 Oct;58(10):2777-2786. [doi: 10.1109/tbme.2011.2155064]

21. Yu L, Yuen PC. Lai J: Ulcer detection in wireless capsule endoscopy images, In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE, 2012 Nov;45-48. 2012 Presented at: 21st International Conference on Pattern Recognition (ICPR 2012); November 11-15, 2012; Tsukuba, Japan.
22. Charisis V, Katsimerou C, Hadjileontiadis L, Liatsos C. Sergiadis GD: Computer-aided capsule endoscopy images evaluation based on color rotation and texture features: An educational tool to physicians, In Proceedings of the 26th IEEE International Symposium On Computer-Based Medical Systems, IEEE, 2013;203-208. 2013 Jun 20 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627789]
23. Eid A, Charisis VS, Hadjileontiadis LJ, Sergiadis GD. A curvelet-based lacunarity approach for ulcer detection from wireless capsule endoscopy images, In Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, IEEE, 2013; 273-278. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627801]
24. Yeh J, Wu T, Tsai W. Bleeding and ulcer detection using wireless capsule endoscopy images. JSEA 2014;07(05):422-432. [doi: 10.4236/jsea.2014.75039]
25. Yuan Y, Wang J, Li B, Meng MQ. Saliency based ulcer detection for wireless capsule endoscopy diagnosis. IEEE Trans Med Imaging 2015 Oct;34(10):2046-2057. [doi: 10.1109/TMI.2015.2418534] [Medline: 25850085]
26. Suman S, Hussin F, Malik A, Ho S, Hilmi I, Leow A, et al. Feature selection and classification of ulcerated lesions using statistical analysis for WCE images. Applied Sciences 2017 Oct 24;7(10):1097. [doi: 10.3390/app7101097]
27. Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys Med Biol 2018 Aug 10;63(16):165001. [doi: 10.1088/1361-6560/aad51c] [Medline: 30033931]
28. Alaskar H, Hussain A, Al-Aseem N, Liatsis P, Al-Jumeily D. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. Sensors (Basel) 2019 Mar 13;19(6):1265-1265 [FREE Full text] [doi: 10.3390/s19061265] [Medline: 30871162]
29. Aoki T, Yamada A, Aoyama K, Saito H, Tsuboi A, Nakada A, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2019 Feb;89(2):357-363.e2. [doi: 10.1016/j.gie.2018.10.027] [Medline: 30670179]
30. Charfi S, El Ansari M. Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy videos, In 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), IEEE, 2017;1-5. 2017 Presented at: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP); May 22-24, 2017; Fez, Morocco. [doi: 10.1109/atsip.2017.8075590]
31. Wang S, Xing Y, Zhang L, Gao H, Zhang H. A systematic evaluation and optimization of automatic detection of ulcers in wireless capsule endoscopy on a large dataset using deep convolutional neural networks. Phys Med Biol 2019 Dec 05;64(23):235014. [doi: 10.1088/1361-6560/ab5086] [Medline: 31645019]
32. Wang S, Xing Y, Zhang L, Gao H, Zhang H. Deep convolutional neural network for ulcer recognition in wireless capsule endoscopy: experimental feasibility and optimization. Comput Math Methods Med 2019;2019:7546215 [FREE Full text] [doi: 10.1155/2019/7546215] [Medline: 31641370]
33. Klang E, Barash Y, Margalit RY, Soffer S, Shimon O, Albshesh A, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. Gastrointest Endosc 2020 Mar;91(3):606-613.e2. [doi: 10.1016/j.gie.2019.11.012] [Medline: 31743689]
34. Kundu AK, Fattah SA, Wahid KA. Multiple Linear Discriminant Models for Extracting Salient Characteristic Patterns in Capsule Endoscopy Images for Multi-Disease Detection. IEEE J Transl Eng Health Med 2020;8:3300111 [FREE Full text] [doi: 10.1109/JTEHM.2020.2964666] [Medline: 32190429]
35. Li B, Meng MQ. Computer-aided detection of bleeding regions for capsule endoscopy images. IEEE Trans Biomed Eng 2009 May;56(4):1032-1039. [doi: 10.1109/TBME.2008.2010526] [Medline: 19174349]
36. Penna B, Tillo T, Grangetto M, Magli E, Olmo G. A technique for blood detection in wireless capsule endoscopy images. 2009 Presented at: 17th European Signal Processing Conference; Aug 24-29, 2009; Glasgow, Scotland.
37. Fu Y, Zhang W, Mandal M, Meng MQ. Computer-aided bleeding detection in WCE video. IEEE J Biomed Health Inform 2014 Mar;18(2):636-642. [doi: 10.1109/JBHI.2013.2257819] [Medline: 24608063]
38. Ghosh T, Bashar S, Alam M, Wahid K, Fattah SA. A statistical feature based novel method to detect bleeding in wireless capsule endoscopy images, IEEE, 2014;1-4. 2014 Presented at: The 17th International Conference on Informatics, Electronics & Vision; May 23-24, 2014; Dhaka, Bangladesh. [doi: 10.1109/iciev.2014.6850777]
39. Sainju S, Bui FM, Wahid KA. Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. J Med Syst 2014 May;38(4):25. [doi: 10.1007/s10916-014-0025-1] [Medline: 24696394]
40. Dilna C, Gopi V. A novel method for bleeding detection in Wireless Capsule Endoscopic images. 2015 Presented at: International Conference on Computing and Network Communications; Dec 16-19, 2015; Trivandrum, India. [doi: 10.1109/coconet.2015.7411289]
41. Ghosh T, Fattah S, Bashar S, Shahnaz C, Wahid K, Zhu WP. An automatic bleeding detection technique in wireless capsule endoscopy from region of interest, In. 2015 Jul 21 Presented at: 2015 IEEE International Conference on Digital Signal Processing; July 21-24, 2015; Singapore. [doi: 10.1109/icdsp.2015.7252090]

42. Mathew M, Gopi VP. Transform based bleeding detection technique for endoscopic images. 2015 Feb 26 Presented at: The 2nd International Conference on Electronics and Communication Systems (ICECS); Feb 26-27, 2015; Coimbatore, India. [doi: 10.1109/ecs.2015.7124882]
43. Xiao Jia, Meng MQ. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:639-642. [doi: 10.1109/EMBC.2016.7590783] [Medline: 28268409]
44. Liu D, Gan T, Rao N, Xing Y, Zheng J, Li S, et al. Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process. Med Image Anal 2016 Aug;32:281-294. [doi: 10.1016/j.media.2016.04.007] [Medline: 27236223]
45. Yuan Y, Li B, Meng MQ. Bleeding Frame and Region Detection in the Wireless Capsule Endoscopy Video. IEEE J Biomed Health Inform 2016 Mar;20(2):624-630. [doi: 10.1109/JBHI.2015.2399502] [Medline: 25675468]
46. Xiao Jia, Meng MQ. Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. Annu Int Conf IEEE Eng Med Biol Soc 2017 Jul;2017:3154-3157. [doi: 10.1109/EMBC.2017.8037526] [Medline: 29060567]
47. Leenhardt R, Vasseur P, Li C, Saurin JC, Rahmi G, Cholet F, CAD-CAP Database Working Group. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointest Endosc 2019 Jan;89(1):189-194. [doi: 10.1016/j.gie.2018.06.036] [Medline: 30017868]
48. Aoki T, Yamada A, Kato Y, Saito H, Tsuboi A, Nakada A, et al. Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. J Gastroenterol Hepatol 2020 Jul;35(7):1196-1200. [doi: 10.1111/jgh.14941] [Medline: 31758717]
49. Tsuboi A, Oka S, Aoyama K, Saito H, Aoki T, Yamada A, et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. Dig Endosc 2020 Mar;32(3):382-390. [doi: 10.1111/den.13507] [Medline: 31392767]
50. Vuik FER, Nieuwenburg SAV, Moen S, Schreuders EH, Oudkerk Pool MD, Peterse EFP, et al. Population-Based Prevalence of Gastrointestinal Abnormalities at Colon Capsule Endoscopy. Clin Gastroenterol Hepatol 2020 Oct 31:S1542-S3565. [doi: 10.1016/j.cgh.2020.10.048] [Medline: 33130189]
51. Murphy B, Winter DC, Kavanagh DO. Small Bowel Gastrointestinal Bleeding Diagnosis and Management-A Narrative Review. Front Surg 2019;6:25 [FREE Full text] [doi: 10.3389/fsurg.2019.00025] [Medline: 31157232]
52. Bang CS, Ahn JY, Kim J, Kim Y, Choi IJ, Shin WG. Establishing Machine Learning Models to Predict Curative Resection in Early Gastric Cancer with Undifferentiated Histology: Development and Usability Study. J Med Internet Res 2021 May 15;23(4):e25053 [FREE Full text] [doi: 10.2196/25053] [Medline: 33856358]
53. Bang CS, Lee JJ, Baik GH. Computer-aided diagnosis of diminutive colorectal polyps in endoscopic images: systematic review and meta-analysis of diagnostic test accuracy. J Med Internet Res 2021 Aug 25;23(8):e29682 [FREE Full text] [doi: 10.2196/29682] [Medline: 34432643]
54. Bang CS, Lee JJ, Baik GH. Artificial intelligence for the prediction of Helicobacter pylori infection in endoscopic images: systematic review and meta-analysis of diagnostic test accuracy. J Med Internet Res 2020 Sep 16;22(9):e21983 [FREE Full text] [doi: 10.2196/21983] [Medline: 32936088]

Abbreviations

AUC: area under the curve
CAD: computer-aided diagnosis
DOR: diagnostic odds ratio
DTA: diagnostic test accuracy
FP: false positive
FN: false negative
HSROC: hierarchical summary receiver operating characteristic
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for a Systematic Review and Meta-analysis
PROSPERO: International Prospective Register of Systematic Reviews
QUADAS-2: the second version of Quality Assessment of Diagnostic Accuracy Studies
TP: true positive
TN: true negative
WCE: wireless capsule endoscopy

Edited by G Eysenbach; submitted 30.08.21; peer-reviewed by SI Seo; comments to author 09.10.21; revised version received 10.10.21; accepted 13.10.21; published 14.12.21.

Please cite as:
Bang CS, Lee JJ, Baik GH
Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-analysis
J Med Internet Res 2021;23(12):e33267
URL: https://www.jmir.org/2021/12/e33267
doi:10.2196/33267
PMID:34904949

Original Paper

How Clinicians Perceive Artificial Intelligence–Assisted Technologies in Diagnostic Decision Making: Mixed Methods Approach

Hyeyoung Hah[1], PhD; Deana Shevit Goldin[2], PhD, DNP

[1]Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

[2]Nicole Wertheim College of Nursing & Health Sciences, Florida International University, Miami, FL, United States

Corresponding Author:
Hyeyoung Hah, PhD
Information Systems and Business Analytics
College of Business
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3053484342
Email: hhah@fiu.edu

Abstract

Background: With the rapid development of artificial intelligence (AI) and related technologies, AI algorithms are being embedded into various health information technologies that assist clinicians in clinical decision making.

Objective: This study aimed to explore how clinicians perceive AI assistance in diagnostic decision making and suggest the paths forward for AI-human teaming for clinical decision making in health care.

Methods: This study used a mixed methods approach, utilizing hierarchical linear modeling and sentiment analysis through natural language understanding techniques.

Results: A total of 114 clinicians participated in online simulation surveys in 2020 and 2021. These clinicians studied family medicine and used AI algorithms to aid in patient diagnosis. Their overall sentiment toward AI-assisted diagnosis was positive and comparable with diagnoses made without the assistance of AI. However, AI-guided decision making was not congruent with the way clinicians typically made decisions in diagnosing illnesses. In a quantitative survey, clinicians reported perceiving current AI assistance as not likely to enhance diagnostic capability and negatively influenced their overall performance (β=–0.421, P=.02). Instead, clinicians' diagnostic capabilities tended to be associated with well-known parameters, such as education, age, and daily habit of technology use on social media platforms.

Conclusions: This study elucidated clinicians' current perceptions and sentiments toward AI-enabled diagnosis. Although the sentiment was positive, the current form of AI assistance may not be linked with efficient decision making, as AI algorithms are not well aligned with subjective human reasoning in clinical diagnosis. Developers and policy makers in health could gather behavioral data from clinicians in various disciplines to help align AI algorithms with the unique subjective patterns of reasoning that humans employ in clinical diagnosis.

(J Med Internet Res 2021;23(12):e33540) doi:10.2196/33540

KEYWORDS

artificial intelligence algorithms; AI; diagnostic capability; virtual care; multilevel modeling; human-AI teaming; natural language understanding

Introduction

Overview

Artificial intelligence (AI) and related technologies are rapidly evolving as part of workplace technology to support clinicians' decision making [1,2]. AI refers to the use of a collection of intelligent technologies in "the science and engineering of intelligent machines and computational part of the ability to achieve goals in the world" [3] or to "model intelligent behavior with minimal human intervention" [4-6]. Its influence has permeated retail, marketing, and human resource management [6]. Specifically in health care, considering the risks of incorrect predictions, the Consumer Technology Association (CTA) has reconceptualized health care AI as "assistive intelligence," described as "a category of AI-enabled software that 'informs' or 'drives' diagnosis or clinical management of a patient," in which "the health care provider makes the ultimate decisions before clinical action is taken" [7]. In both clinical and administrative areas of health care organizations, AI algorithms are increasingly embedded in health information technologies (HITs) to assist with clinical functions such as data monitoring, clinical research, diagnostics, and support compliance for billing and administrative tasks [8,9]. AI technology in health care is therefore expected to streamline clinicians' clinical and administrative decision making by providing prompt data analyses and the necessary recommendations to make health care more efficient and cost-effective.

Particularly within the clinical side of health care organizations, clinical decision making is based on subjective and objective patient information and other influencing variables. Clinical decision making generally refers to the process of making a choice among options aimed at diagnosis, intervention, interaction, and evaluation within a context, in which numerous interactions exist among stakeholders, background knowledge, and social and technological factors [10]. In other words, clinicians' decision making regarding patient health outcomes or health care diagnoses is largely influenced by numerous factors such as the idiosyncratic characteristics of patients and clinicians, consultation environments, and technology used [11]. In this information-rich context, the ability to access and manage appropriate and accurate information is critical for clinicians to make decisions on the patient's behalf. As reflected in the CTA definition, AI algorithms under these circumstances need to complement human cognitive processing of electronic medical records, multimedia images, or laboratory results. Once successfully incorporated, AI assistance has been shown to help clinicians reduce adverse events outside of the intensive care unit by 44% [12].

Thus far, views have been mixed on how AI may "assist" or "team up with" clinicians through the medical assessment and diagnosis processes. On the one hand, AI allows clinicians to freely follow their own paths of inquiry through patient investigations and select from a database of questions on patient history, physical examinations, and laboratory tests to make patient-specific diagnostic decisions [13]. On the other hand, training AI with a collection of data from the clinicians' own subjective diagnoses of various patient cases may not be ideal for fully utilizing AI techniques [14]. Indeed, inconsistent performance has also been reported [15]. Given that 83% of health care organizations have a strategy for AI investment and deployment in the coming years [16], particularly for electronic health record management and diagnosis [17], health care practice and training will follow such trends. However, as core users of AI, incumbent health care stakeholders in the current health care system have differential levels of technological readiness; therefore, clinicians' existing ability to incorporate analytic results from AI may hamper the leveraging of the full potential of AI. In other words, care providers' attitudes toward and positive experiences with the assistance offered by AI-enabled technologies in the process of clinical diagnosis can help health care organizations reap the benefits of AI investments by enhancing consistency, quality of care, and reducing costs [18].

However, research on whether and how clinicians assess AI assistance in decision-making processes remains limited. Thus far, in health care AI literature, clinicians' clinical task performance is known to be influenced by clinical task types and evidence-based standards that support AI's data structure and clinical integration capacity within health care institutions [19]. One stream of AI literature, which has focused on the types of tasks and medical diagnosis, claims that AI algorithms enhance human clinicians' capabilities while improving efficiency in incident reporting and reducing adverse events [20,21]. However, the quality of health care diagnostic tasks between clinicians and AI continues to be characterized by broad discrepancies [22-24]. According to literature on AI's data structure, health information management professionals' ability to improve coded data quality and data patterns is necessary to ensure the optimal adaptation of teaming with AI [25]. Moreover, per the literature on the clinical integration of AI, clinicians' perceptions should be recognized to enhance the clinical diagnosis by AI [21,26-29].

In summary, our literature review revealed that, although advanced AI algorithms have the potential to enhance the quality and efficiency of health care, the critical users of AI algorithms are clinicians whose roles are to understand and communicate with AI. Nonetheless, there has been limited focus in AI health care research on clinicians' attitudes toward AI assistance in actual diagnostic decision making, which may be due to the infancy of such AI-assisted tools, lack of trust in AI from health care stakeholders, and potential health-related risks. Therefore, there is reasonable urgency to examine clinicians' attitudes and sentiments toward AI assistance in clinical decision making, to address the gaps in the existing body of knowledge.

Background and Theory

Medical Diagnostic Knowledge Theory

Clinicians' decision making includes diagnoses or high-complexity decisions across medical specialties. A diagnosis is viewed as an iterative process of "task categorization" by decomposing patients' health symptoms into different task classes and matching the given conditions with the predefined disease categories based on their respective hypotheses [30]. Clinicians are expected to predict and determine the course of action based on their knowledge and

experience and by utilizing information on the features of the focal clinical situations. The theory of clinical diagnosis revolves around 2 important concepts: clinical knowledge and clinical reasoning. The former refers to the fundamental base of knowledge, and it can be subdivided into 3 categories: conceptual, strategic, and conditional knowledge [31]. The latter refers to a holistic process of hypothesis generation, pattern recognition, context formulation, diagnostic test interpretation, differential diagnosis, and diagnostic verification, all of which provide both the language and methods of clinical problem solving [31]. Once clinicians have obtained clinical knowledge, they then organize it using mental representations, medical scripts, or clinical examples, followed by cognitive processes to translate medical information into testable hypotheses in each context and evaluate their own clinical reasoning for clinical diagnosis [32]. Specifically, the conceptual framework of a clinical diagnosis includes the "hypothetico-deductive model" (ie, generation of multiple competing hypotheses from preliminary information provided by patients), decision analysis, pattern recognition, and intuition [33]. During this process, challenges exist as to how clinicians match patients' cases with known patterns and focus on complex patient information among many treatment options [33]. In summary, a typical clinical decision-making process consists of the speedy processing of the situational features, assessment of the relevant hypotheses, and investigations and treatments in a sequence [30].

Such clinical reasoning by individual clinicians cannot be fully supported by AI algorithms' generalized task categorization, pattern recognition, and matching [34]. The current level of AI performs well for clinical knowledge creation by simple interpretations of medical images, slides, and graphs, as well as detection of complex relational time-series patterns within data sets [35,36]. AI algorithms for clinical reasoning (ie, IBM Watson, chatbots to smartphone apps) are believed to function as decision support tools for making diagnoses, providing patient consultation, and detecting certain medical symptoms [26].

In his book *Deep Medicine*, Eric Topol [37] highlighted the role of AI in clinical diagnosis as augmented intelligence: Some clinical knowledge formation via feature selection, task categorization, and pattern recognition can be aided by AI for clinicians to perform clinical reasoning by making a clinical diagnosis. A clinical diagnosis is a high-complexity decision that needs to be based on information inquiry using idiosyncratic patient data and threshold-based decision rules. Without clearcut threshold rules for decisions, AI may not fully process and analyze idiosyncratic patient cases and inaccurate descriptions from patients; it may likewise fail to incorporate the complexity of the diagnosis and predict patients' adverse conditions [35]. Therefore, the role of AI in clinical reasoning or clinical diagnosis is expected to be more assistive, and clinicians should communicate well with AI to prevent any adverse effects of clinical reasoning on patients' health outcomes.

Artificial Intelligence Technology Use in Practice and in Simulation

In practice and in simulation, clinicians are users or collaborators of such "assistive" AI, contingent upon the extent and scope of such technologies that are defined within the context. In other words, AI can refer to algorithms that are embedded in existing HITs or holistic technological artifacts to be newly implemented as standalone software. As such, the use of AI is no longer limited to specific HITs in practice. Clinicians may have unintentionally encountered various AI technologies through system interfaces and embedded AI decision-making logic within systems used in medicine. Owing to such mixed definitions of AI in health care [38], the effects of AI-assisted diagnostic performance have been mixed in prior studies. Positive performance is expected when clinicians trust and expect the performance of AI [21], whereas information overload driven by AI algorithms can cause a loss of situational awareness, thereby negatively affecting clinicians' decision-making abilities [35]. In addition, clinicians are perceived as more trusted by patients [26] and that AI algorithms may not be effective in diagnosing unique cases. Thus, clinicians need to diagnose and communicate with patients using AI, at least for a while. The manner through which clinicians make clinical diagnoses with the support of AI technology and perceive such a diagnostic decision merits further investigation.

Unlike these practical constraints, observations and evaluations of clinicians' health care diagnostic behaviors have been methods of inducing behavioral changes in safer simulation contexts. As what is learned throughout clinical health care training has been associated with what is likely to be carried out in clinical practice [35] and similar technologies span across practice and simulation contexts [39], clinicians' AI attitudes and performance have been studied and predicted from their use of simulated AI [40]. Thus, one can examine the way clinicians have used and familiarized themselves with AI in simulation and extrapolate their behaviors in practice.

Given that clinicians' behaviors and readiness to use this technology can be predicted from simulation experiences [39], we turned to a context of clinical diagnosis simulations—whereby clinicians have familiarized themselves with AI—and evaluated their decisions in a safe and controlled context. Subsequently, we examined whether and how clinicians perceive AI assistance and how their perceptions may differ from other non-AI-based diagnostic situations.

Taken together, in simulation contexts, this study explored users' detailed experiences with AI-enabled patient diagnosis and examined the effect of AI assistance on diagnostic performance. Thus, the following research questions were formulated to shed light on clinicians' attitudes and behavioral characteristics regarding AI assistance in patient diagnosis:

- Research question 1. During patient diagnosis, what are clinicians' sentiments toward AI assistance, and how do their sentiments differ from other non-AI-based diagnostic situations?
- Research question 2. How does current AI assistance affect clinicians' perceptions of enhancing diagnosis and future care task performance?

Methods

Survey Procedure

To this end, our target population was clinicians who have experience with patient diagnosis encounters using AI-based diagnostic technology to understand clinicians' perceptions of AI-assisted diagnosis in a controlled and safe context where patient care was not compromised by potentially incorrect AI algorithms. We recruited clinicians who met the abovementioned requirement and used AI-based diagnostic technology using live patient simulations in a nursing simulation lab. We accessed 3 cohorts of family nurse practitioner (FNP) students who experienced 3 patient diagnosis simulations in the lab: encounters with live standard patients, encounters with AI assistance in diagnosing patients using multimedia patient information, and encounters with patient simulators (ie, lifelike mannequin patients). Each qualified participant was then incentivized by US $5 Amazon gift cards for their completion and the quality of their responses. Consequently, 144 clinicians were selected and invited to participate in the online simulation surveys.

In the simulation lab at the College of Nursing, AI-enabled diagnosis technology has been used in on-ground lessons, in which faculty and students collaborate to complete virtual patient cases and in home-based diagnostic decision making [13,41,42]. In our university, after an AI system developer implemented the focal AI technology and trained the graduate nursing faculty on the technology and after the successful go-live events, this technology was subsequently integrated as part of the clinical simulations for the students. The software is based on data from hundreds of actual patients compiled by experts and AI and based on physiology algorithms [13]. The use of this system is particularly emphasized in the final year of the program to promote skills building and practice in clinical diagnosis.

This interactive AI incorporates intelligence from both humans and AI physiology algorithms such that evaluators on the other end of the system can recognize any patterns demonstrated by those specific users in the process of clinical diagnosis. Clinicians are known to experience some technical features of AI, such as search ability, knowledge expression function, reasoning ability, abstraction ability, speech recognition ability, and ability to process fuzzy clinical information [13]. In our context, reasoning ability was embedded by FNP faculty in the AI-enabled diagnosis technology so that the participants could make diagnostic decisions with AI assistance and obtain feedback after completing each patient case.

Survey Instruments

Our online simulation surveys consisted of 2 parts: survey instruments and open-ended simulation questions. First, for the survey instruments, each clinician reported their perception of AI assistance and perceived performance of patient diagnosis and overall clinical tasks using a 7-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7). Table 1 presents the key survey items sourced from existing information systems (IS) literature.

As shown in Table 1, our study had 2 dependent variables. One was diagnostic performance, which we defined as the ability to provide health care consultations and diagnose health-related issues properly both in person and via online or telehealth platforms [3]. We contextualized and operationalized clinicians' diagnostic capabilities in a virtual context (mean 5.39, SD 0.13) [30]. The other dependent variable was clinical task performance, the survey items for which were adapted from the IS literature (mean 5.48, SD 0.12) [43].

We included control variables that accounted for personal technological traits and demographics. On the one hand, personal technological traits were measured via individual levels of technological advancement, such as personal innovativeness [44], technological habits [45], and computer literacy [46]. In the IS literature, personal innovativeness and computer literacy have been salient in explaining the technology adoption behaviors of individual users. Here, we defined personal innovativeness as "the willingness of an individual to try out any new information technology" [1], whereas computer literacy is "a judgement of clinicians' capability to use a computer" [6]. We also included technology habits to control for the participants' potential automatic reactions to the use of AI-enabled diagnosis technology that includes multimedia information and similar AI algorithms on social media platforms [47]. Lastly, participants' demographic characteristics, such as self-identified gender, age, income, education, and occupation, were included in the survey to control for potential confounding effects.

Next, in the part listing open-ended simulation questions, each clinician was asked to describe their experience with patient diagnosis under 3 different diagnostic modalities: diagnosing a live patient, diagnosing a human-like mannequin, and AI-based diagnostic simulations using AI assistance. In the 3 simulation prompts, participants were asked to recall their completed patient cases during the semester and write comments using either keywords or key phrases. For example, particularly for the AI case, the scenario prompt reads as given in Textbox 1. After reading this prompt, participants described "favorite" as well as "least favorite" diagnosis experiences using keywords or key phrases in 2 open-ended questions. Each clinician's 6 diagnostic encounters were recorded as textual narratives, along with the 3 different simulation contexts of the patient diagnoses.

Table 1. Key survey items.

Variables	Survey items[a]	References
Dependent variables		
Diagnostic performance	• OO[b] system allows me to carefully evaluate the health condition of the patient. • OO system allows me to thoroughly assess the health condition of the patient. • OO system allows me to accurately evaluate the patient's health condition. • OO system allows me to think critically during the simulation experience.	[30]
Clinical task performance	• I believe that the use of OO system can increase my overall performance. • I believe that the use of OO system can increase my effectiveness with the care tasks when working with live patients in the future. • With the use of OO system, I believe I can work more efficiently for managing care tasks when working with live patients in the future. • I believe that the use of OO system can increase the quality of care. • I believe that the use of OO system can decrease error rates in communication and information sharing with other care members in the future. • I believe that the use of OO system will help me understand what I have learned.	[43]
Independent variable		
AI[c] assistance	• Clinicians' experience of AI-assisted diagnostic simulation (binary: 1, with AI assistance; 0, live patient encounter with no AI assistance)	
Key control variables		
Personal technology trait: technology habit	• The use of social media has become a habit for me at work. • I am addicted to using social media at work. • I must use social media at work. • Using social media has become natural to me at work	[44]
Personal technology trait: personal innovativeness	• If I heard about a new information technology, I would look for ways to experiment with it. • In general, I am hesitant to try out new information technologies. • Among my peers, I am usually the first to try out new information technologies. • I like to experiment with new information technologies.	[45]
Personal technology trait: computer literacy	• I could complete the health care task using health information technology if there was no one around to tell me what to do as I go • I could complete the health care task using health information technology if I had just the built-in help menu for assistance. • I could complete the health care task using health information technology if someone showed me how to do it first. • I could complete the health care task using health information if I had used similar apps before this one to do the same job.	[46]

[a]Each item uses a seven-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7).

[b]OO system refers to artificial intelligence–enabled diagnosis technology in our research setting.

[c]AI: artificial intelligence.

Textbox 1. Example scenario of artificial intelligence (AI)–based patient diagnosis.

> You read a case description about a 50-year-old patient on an AI-based diagnosis system. His complaints are about back pain. "My back has been hurting for around five days. I bent over to pick something up in my print shop and I had this severe pain in my back. It hurts so much so it is hard to stand up. The pain is on and off throughout the day—maybe four times, for half an hour each, especially when I am walking around. It is mostly a dull aching pain."

Study Design

Using clinicians' quantitative and qualitative responses from our online simulation surveys, we utilized a mixed method to analyze our mixed data. According to Creswell [48] and Creswell and Clark [49], mixed methods research "incorporates qualitative and quantitative data to solve complex research questions and hypotheses, and it is suitable for explaining what (ie, estimating overall trends in participant behaviors within the population) and obtaining an in-depth understanding of why (ie, specific individual behaviors in the subsamples)." This method is particularly useful for exploring context-specific variables and understanding beyond the importance of research variables by delving into the underlying mechanisms using causal models.

Mixed methods research includes sequential and concurrent designs such that qualitative and quantitative data can be collected sequentially in the former, whereas both types of data are obtained at the same time in the latter [50]. Each category has a specific design typology based on the emphasis on the importance of qualitative or quantitative data, the analysis process, and theoretical emphasis (see Castro et al [50] for an in-depth discussion on designs of mixed methods research). We applied a concurrent triangulation design, because both quantitative and qualitative data were collected concurrently, and such data were used to accurately describe and examine clinicians' experiences and sentiments toward AI assistance in clinical decision making [50].

Statistical Analysis

We utilized qualitative and quantitative data using natural language understanding (NLU) and hierarchical linear modeling (HLM). Regarding the first research question—capturing clinicians' respective sentiments in 3 cases of patient diagnosis simulations—we analyzed textual narratives to understand what clinicians perceived about the diagnosis process in teaming with AI and compared it with non-AI-involved diagnostic situations using NLU. Textual comments consisted of a single sentence or a small number of keywords. To incorporate such data characteristics, we deemed NLU, as a subfield of computer science, an adequate method based on its explicit focus on the use of computational techniques to learn, understand, and produce human language content, and many information technology firms, such as Microsoft, Google, and IBM, have developed NLU platforms and algorithms [51,52]. We utilized IBM Watson Natural Language Understanding-77 for its well-known performance and cross-evaluation results. Understanding each participant's language content was more suitable than traditional text mining analytics [52-54].

Next, for quantitative data analysis, we implemented HLM estimation techniques to answer the second research question—measuring the effect of the current form of AI assistance on clinicians' diagnostic decision making and care task performance. In our data, each clinician's clinical diagnosis experience was nested within the program. In other words, under the same graduate program, each participant was exposed to the same sets of patient diagnosis simulations with 3 modalities and reported their responses with 3 diagnosis simulations repeatedly. This context could engender statistical dependency among the responses [55]. Moreover, in the research model, the outcome variable (ie, the clinical diagnosis) was at the individual level, and the AI assistance variable was at the group level (or graduate nursing program), leading to a multilevel research program in the same research model [21]. To mitigate such dependency and estimate this multilevel model, we used the STATA 16.1 (StataCorp, College Station, TX) xtmixed procedure to carry out generalized linear modeling using the restricted maximum likelihood or residual maximum likelihood estimation. The xtmixed procedure fits linear mixed models that include fixed (or standard regression coefficients) and random effects that are not estimated directly but by variance component. The default covariance structure is independent covariance. We also specified other covariance structures (eg, exchangeable and unstructured) to validate our mixed model results (see [56] for an in-depth discussion of mixed models and covariance structures).

Results

We explored clinicians' perceptions and performance prospects toward AI assistance in patient diagnosis research. These relations and sentiments are evidenced by the results of generalized linear models as well as qualitative text analysis with direct quotations from the health care worker sample.

User Statistics

A total of 114 clinicians completed our online surveys during the 2020-2021 study period (response rate: 114/144, 79.2%). In summary, 66.7% (76/114) of the participants were between 26 years old and 40 years old, 49.1% (56/114) were white, and 84.2% (96/114) identified as female. Additionally, 89.5% (102/114) worked either full time or part time in hospitals or clinics. In terms of education, 24.6% (28/114) had a graduate degree, whereas all the participants had obtained a bachelor's degree in nursing with prior clinical experience before joining the graduate program, as shown in Table 2.

Table 2. Respondent demographics (N=114).

Characteristics	Results, n (%)
Self-identified gender	
Male	16 (14.0)
Female	96 (84.2)
Not disclosed	2 (1.8)
Age (years)	
18-25	10 (8.8)
26-40	76 (66.7)
41-55	26 (22.8)
56-65	2 (1.8)
Income (US $)	
25,000-49,999	20 (17.5)
50,000-74,999	32 (28.1)
75,000-99,999	30 (26.3)
≥100,000	4 (3.5)
Prefer not to answer	28 (24.6)
Education	
Bachelor's degree	82 (71.9)
Master's degree	20 (17.5)
PhD	8 (7.0)
Others	4 (3.5)
Occupation	
Working full time	54 (47.4)
Working part time	48 (42.1)
Unemployed	6 (5.3)
Other	6 (5.3)
Race	
African American	22 (19.3)
Asian	8 (7.0)
Native Hawaiian or Pacific Islander	2 (1.8)
White	56 (49.1)
Other	20 (17.5)
Prefer not to answer	6 (5.3)
Urban/rural area	
Urban	98 (86.0)
Rural	10 (8.8)
Other	6 (5.3)

Research Question 1: Sentiment Analysis Results

To identify clinicians' sentiments toward AI assistance, we compared the clinicians' perceptions regarding AI-assisted diagnosis and non-AI diagnosis contexts. Table 3 presents some narrative examples in our data set whereby the clinicians described what they liked and disliked about the patient diagnosis process with the help of AI-aided technology, with a live human patient, and with a human-like mannequin/patient simulator.

The results of the NLU analyses are reported in Tables 4, 5, and 6. We found that clinicians perceived their patient diagnosis experience differently across the 3 simulation cases. The positive

sentiment scores for the diagnosis process were 0.99 for the live patient, 0.92 for teaming up with AI assistance, and 0.41 for the patient simulator. In contrast, the negative sentiment score was the highest with the patient simulator (sentiment score = –1), followed by the live patient (sentiment score = –0.97) and AI assistance (sentiment score = –0.87). Specifically, our respondents perceived diagnostic simulations with AI technology less negatively and more positively compared with diagnosing a live human patient. Table 4 reports each positive and negative sentiment score for all 3 cases, with scores ranging from –1 to 1.

Tables 5 and 6 present the most relevant keywords from the textual narratives for the 3 simulation cases. The respondents perceived AI-based diagnostic simulations positively for the following reasons: “convenient access,” “thorough assessment skills,” “student interaction convenience,” “interactive learning rationale,” “good learning opportunities” and “[a] vast range of questions.” At the same time, they also perceived the AI-based diagnosis simulation to contain a “long system,” “strict sensitive clicking,” “differential diagnosis,” “large learning curve,” and “technical difficulties,” and they found it to be a “hard system.” These keywords indicated that the current form of AI assistance may lead to differential diagnosis logic relative to clinicians’ own logic and hypotheses, and technical difficulties could cause users to be averse to the technology.

Table 3. Some narrative examples in our data set, based on 65 recorded instances.

Narrative valence	AI[a] assistance context	Non-AI assistance context	
	Diagnosis experience with AI assistance	Diagnosis experience with live patients	Diagnosis experience with HPS[b]
Positive comments	“I don’t feel like I am under pressure and can do it at my own pace”	“interaction, being able to gauge the patient, reading facial expressions, immediate feedback”	“Very realistic, lifelike scenarios”
Negative comments	“not having an orientation on how to work the system (first time user)”	“I'm not an actor, and it felt like acting; immediate feedback”	“being watched through a one-way mirror”

[a]AI: artificial intelligence.

[b]HPS: human patient simulator. It is worth nothing that the clinicians recorded their retrospective experience with the HPS, as it was not used in the graduate program.

Table 4. Clinicians’ emotions with patient diagnosis under 3 different scenarios, based on 65 recorded instances.

Sentiment	AI[a] assistance context	Non-AI assistance context	
	Diagnosis experience with AI assistance	Diagnosis experience with live patients	Diagnosis experience with HPS[b]
Positive sentiment[c]	0.92	0.99	0.41
Negative sentiment[c]	–0.87	–0.97	–1

[a]AI: artificial intelligence.

[b]HPS: human patient simulator. It is worth nothing that the clinicians recorded their retrospective experience with the HPS, as it was not used in the graduate program.

[c]The sentiment score ranged from –1 (negative) to 1 (positive).

Table 5. Extracted keywords from clinicians' positive textual narratives, based on 65 recorded instances.

Keyword	Relevance[a]
AI[b] assistance context: diagnosis experience with AI assistance	
Convenient access	0.976987
Thorough assessment skills	0.641589
Good practice student interaction convenience	0.62027
Interactive learning rationales	0.612274
Good learning opportunities	0.599706
Questions	0.559626
Vast list of questions	0.542493
Question banks	0.537307
Times	0.534484
Issues	0.518757
Convenience	0.515202
Scenario	0.514838
History	0.513546
Patient data	0.512584
Plan	0.507775
Pressure	0.5077
Diagnoses	0.507442
Students	0.507159
Ease	0.506215
Choices	0.505864
Non-AI assistance context: diagnosis experience with live patients	
Fast convenient real experience	0.89935
Fast convenience	0.706168
Real-life situation	0.68436
High quality	0.68049
Telehealth: convenient fast access	0.664928
Real world	0.602068
Real life	0.600372
Live patient	0.581455
Physical actions convenience	0.581031
Video interactions	0.561663
Convenience	0.546701
Best learning experience	0.540142
Challenging open-ended questions	0.536408
Live experience	0.534688
Positive feedback	0.533209
Patients	0.532791
Facial expressions	0.529643
Experience	0.528151
Common complaint	0.526369

Keyword	Relevance[a]
Non-AI assistance context: diagnosis experience with HPS[c]	
Good practice	0.721493
Real patient	0.696931
Less fear	0.695137
Convenient reinforcement of learning	0.633235
Better interaction	0.60778
Real world	0.591024
New things	0.58642
Future trend	0.58135
Clear case	0.581078
Convenience	0.554921
Patient simulators	0.543254
Point	0.527845
Scenarios	0.519927
Mistake	0.518499
Practice maneuvers	0.513818
Patients	0.512968
Experience	0.512967
Ease	0.510172
Survey	0.507779
Person	0.507779

[a]Relevance scores range from 0 to 1, reflecting that higher values indicate greater relevance.

[b]AI: artificial intelligence.

[c]HPS: human patient simulator. It is worth nothing that the clinicians recorded their retrospective experience with the HPS, as it was not used in the graduate program.

Table 6. Extracted keywords from clinicians' negative textual narratives, based on 65 recorded instances.

Keyword	Relevance[a]
AI[b] assistance context: diagnosis experience with AI assistance	
First-time user	0.837534
Long system	0.779516
Strict sensitive clicking	0.634627
Differential diagnosis	0.62574
Large learning curve	0.610383
Technical difficulties	0.584259
Hard system	0.573913
Results of x-rays and CT[c]	0.567441
Sound doesn't work	0.552959
Cases	0.545605
Next part	0.542253
Complex	0.538614
Times	0.531118
Area	0.522729
Orientation	0.520495
List	0.518181
Real patient	0.513188
User	0.512912
Work	0.510401
Things	0.509941
Non-AI assistance context: diagnosis experience with live patients	
Strict testing environment	0.7036
Face interaction complex	0.65795
Limited time	0.65711
Short time	0.60814
Constant need	0.59631
Real clinic patients	0.59044
Physical examination (PE)	0.5563
Sound effect	0.55628
Feeling of self-doubt	0.54479
Accurate data	0.54182
Assess patient	0.5376
Patient expresses lack of physical exam	0.53226
Feedback	0.53139
Interaction	0.52951
Actor	0.5278
Minutes	0.51986
PE doesn't correlate	0.51702
Quality distractions	0.51557
Unreliability	0.51449

Keyword	Relevance[a]
Client	0.51445
Non-AI assistance context: diagnosis experience with HPS[d]	
Human experience	0.71766
Physical examination maneuvers	0.65952
Lack of feelings response	0.55535
Live patient additional questions	0.5476
Lab values	0.53833
Unrealistic prefer	0.53269
Actual patient experiences	0.52713
Expressions	0.52625
Immediate feedback	0.51753
HPS encounters	0.51753
Scenario	0.51542
Simulators	0.51339
Student	0.51188
Real patient	0.51112
Assessment	0.50641
Reaction	0.50641
Survey	0.50641
Realistic interaction	0.49813
Sounds effects	0.48723
One-way mirror	0.48654

[a]Relevance scores range from 0 to 1, reflecting that higher values indicate greater relevance.

[b]AI: artificial intelligence.

[c]CT: computed tomography.

[d]HPS: human patient simulator.

Quantitative Methods: Results of Mixed Models

Table 7 presents our findings for the second research question—the effect of AI assistance on clinicians' clinical diagnosis and care performance from multilevel mixed effects models. Models 1 and 2 report the results for the 2 dependent variables of diagnostic performance and clinical task performance, respectively, where the independent variable is the experience with AI assistance (binary). In terms of individual-level covariates, social media use, personal innovativeness, and computer literacy were included as personal technological traits, in addition to demographic covariates. The HLM results were compared with the baseline ordinary least squares model with clustered standard errors. The effect of AI assistance was not statistically significant in explaining the variation in enhanced clinical diagnosis. Instead, education and age, which were related to clinicians' overall practical experience, were positively associated with clinical diagnosis. The mixed model results were qualitatively identical, along with the different covariance structures.

Table 7. Results from hierarchical linear modeling (N=114 observations).

Variables	Model 1[a]				Model 2[b]			
	OLS[c] (clustered SE[d,e])	*P* values	Mixed model[f,g]	*P* values	OLS (clustered SE[d,h])	*P* values	Mixed model[f,g]	*P* values
Constant	2.162 (1.013)	.04	2.162 (1.851)	.24	4.278 (0.898)	<.001	4.278 (1.462)	.003
AI assistance	–0.105 (0.185)	.57	–0.105 (0.167)	.53	–0.421 (0.192)	.03	–0.421 (0.175)	.02
Personal technological traits								
Technology habit	0.232 (0.104)	.03	0.232 (0.137)	.09	0.244 (0.0803)	.004	0.244 (0.108)	.02
Personal innovativeness	–0.227 (0.202)	.27	–0.227 (0.197)	.25	–0.0234 (0.157)	.88	–0.0234 (0.155)	.89
Computer literacy	–0.161 (0.181)	.38	–0.161 (0.157)	.31	–0.257 (0.111)	.02	–0.257 (0.124)	.04
Control variables								
Female gender	–0.202 (0.615)	.74	–0.202 (0.575)	.73	0.0792 (0.495)	.87	0.0792 (0.454)	.86
Age: 18-25 years	1.885 (0.892)	.04	1.885 (1.473)	.20	–0.910 (0.825)	.28	–0.910 (1.163)	.43
Age: 26-40 years	2.339 (0.782)	.004	2.339 (1.294)	.07	–0.236 (0.704_	.74	–0.236 (1.021)	.82
Age: 41-55 years	2.428 (0.802)	.004	2.428 (1.321)	.07	0.102 (0.683)	.88	0.102 (1.042)	.92
Race: African American	–0.0232 (0.540)	.97	–0.0232 (0.842)	.98	0.00592 (0.615)	.99	0.00592 (0.664)	.99
Race: Asian	–0.419 (0.687)	.55	–0.419 (1.106)	.71	–0.0125 (0.745)	.99	–0.0125 (0.873)	.99
Race: Native Hawaiian/Pacific Islander	1.067 (0.672)	.12	1.067 (1.514)	.48	0.708 (0.724)	.33	0.708 (1.195)	.55
Race: White	–0.0132 (0.417)	.98	–0.0132 (0.780)	.99	0.113 (0.545)	.84	0.113 (0.615)	.85
Race: Other	0.209 (0.631)	.74	0.209 (0.826)	.80	0.644 (0.614)	.30	0.644 (0.652)	.32
Education: Bachelor's degree	1.880 (0.600)	.003	1.880 (1.044)	.07	1.584 (0.495)	.002	1.584 (0.824)	.06
Education: Master's degree	1.586 (0.738)	.04	1.586 (1.128)	.16	1.014 (0.586)	.09	1.014 (0.890)	.25
Education: PhD	0.380 (0.935)	.69	0.380 (1.245)	.76	–0.158 (0.764)	.84	–0.158 (0.982)	.87
Occupational status: working full time	–0.345 (0.413)	.41	–0.345 (0.790)	.66	–0.128 (0.439)	.77	–0.128 (0.623)	.84
Occupational status: working part time	0.332 (0.425)	.44	0.332 (0.801)	.68	0.326 (0.448)	.47	0.326 (0.632)	.61
Occupational status: unemployed	–0.760 (0.485)	.12	–0.760 (1.031)	.46	–0.719 (0.698)	.31	–0.719 (0.813)	.38
Urban	–0.451 (0.433)	.30	–0.451 (0.512)	.38	–0.472 (0.409)	.25	–0.472 (0.404)	.24

[a]Dependent variable: diagnostic performance.

[b]Dependent variable: clinical task performance.

[c]OLS: ordinary least squares.

[d]Robust standard errors are clustered by each participant.

[e]R^2=0.347.

[f]57 groups (clusters).

[g]Variance structures were specified using unstructured, identify, and exchangeable, respectively, and results qualitatively remained the same.

[h]R^2=0.412.

Discussion

Principal Findings

This mixed methods study aimed to explore the status of AI-assisted decision-making patterns among clinicians and gain a detailed understanding of how this novel method of AI-human collaboration and decision making could progress in the future. The results from the qualitative methods showed that clinicians described the diagnostic process with the support of AI as more positive compared with encountering a live patient on their own. Nonetheless, the respondents' keywords revealed that AI assistance impeded clinicians from formulating their own subjective diagnoses based on their clinical reasoning and that

the main complaints among clinicians related to some of the steps in the AI algorithms. Moreover, our quantitative results showed that clinicians' perceptions of their clinical diagnostic capability neither indicated the current level of AI assistance nor enhanced their care task performance. The participants believed that their *ex ante* quality and capability, such as education, age, and daily technology habits, were more relevant in enhancing their care task performance.

Comparison With Prior Work

We expected our research to make 2 important contributions to the existing IS and health care literature. First, our findings from clinicians' keywords in their textual comments demonstrated that clinicians perceived AI assistance positively. However, the current AI interface may not align with their clinical reasoning process, and therefore, such AI interface issues can negatively affect clinicians' perception of whether AI can collaborate with them as a team member. Clinician keywords such as "long system," "strict sensitive clicking," "differential diagnosis," "large learning curve," "technical difficulties," and "hard system" demonstrated human-computer interaction (HCI) issues. In other words, this phenomenon can be linked to the topic of the user interface and explainable or understandable case scenarios in AI research. Studies on HCI have typically focused on the effect of technology on users by considering principles, guidelines, and strategies for designing and interacting with user interfaces [57]. The importance of the design aspect of AI (or AI from an HCI perspective) is more pronounced in health care because active involvement of stakeholders in the design stage can promote appropriation and sensemaking of the focal technology and increase the benefits of AI implementation [58,59]. Studies on machine learning–human interaction have emphasized "human-centeredness" in the use of AI such that humans and machines can integrate or work together as a team [60]. To do so, AI must be explainable, comprehensible, useful, and usable for clinicians in the use scenarios, both in practice and in simulation.

However, some challenges exist. First, the incorporation of fast-moving machine learning techniques into common user experience designs falls under restrictions in environment, law, and regulations in real-world scenarios [57]. Second, there are unintended consequences of AI-generated automated inferences and functions under uncertainty [61] so that, in cases where scenarios yield false positives or false negatives, AI-based decision making negatively affects clinicians' care task performance and patient health outcomes [57]. From the perspective of clinicians, case scenarios or process logic may be the primary interface faced by clinicians under the current development of AI algorithms. However, we also showed that the complexity of AI algorithms may disrupt clinicians' patient diagnostic decisions because clinicians may not understand how to access the complete functions and capabilities of the AI [36]. Thus, based on the previous literature and current evidence in this paper, our results call for more attention to HCI issues by actively involving the end users in the system development procedures and providing them with adequate education to co-create effective decision-making scenarios with AI assistance.

For another, we empirically quantified the effect of AI assistance on clinicians' decision making in a nomological network and explored whether it can serve as a factor to enhance clinical diagnostic capability and overall health task performance. Our results demonstrated that, although clinicians interacted with AI algorithms in safer simulations, the effect of AI assistance was either negative or nonexistent in the clinicians' diagnostic decisions. Instead, the clinicians' *ex ante* personal traits, that is education and age, are positively associated with enhanced outcomes. This finding corresponds with the existing understanding that years of training, professional background, and educational background affect diagnostic performance [62]. Moreover, we found that AI assistance was negatively associated with the clinicians' perceptions of task performance. This might be due to frontline health care providers' trust in AI [14] or implementation issues [63] in health care. It will be worthwhile to revisit our research model to identify AI-specific factors and test downstream effects on clinicians' AI use performance in future research.

Furthermore, we found that the clinicians' *ex ante* technological traits were statistically associated with clinical decision making. First, we found that technological habits were positively linked to clinical diagnostic and health care task performance. In this study, we viewed a technological habit as a social media habit. Clinicians have used social media technologies that may embed AI algorithms and techniques in nonhealth care contexts [64]. It is likely that processing and accessing multimedia-based information daily may help clinicians manage multimedia-based patient information on AI platforms. Second, we showed that computer literacy was negatively associated with overall health care task performance with AI assistance. In health care, clinicians' levels of literacy vary across contexts such as information [65], health [66], YouTube [67], informatics tools [68], and computers [69]. In particular, the concept of computer literacy is broad and encompasses hardware and software [69], informatics tools (eg, decision support systems), handheld devices [68], computerized statistical analysis, databases, presentation graphics, spreadsheet applications, and bibliographic database searches for evidence-based practice [70]. Such measures of literacy are related to knowledge about focal tools or interpreting health information in health care contexts. Meanwhile, a renewed concept of computer literacy has emerged, known as "digital literacy," or the combined knowledge, skills, and competencies necessary for thriving in a technology-saturated culture [71]; it encompasses various forms of literacy via visual, electronic, and digital forms of expression and communication [72]. The use of 3-dimensional virtual images in operations or virtual reality goggles in pathology laboratories [73] or the matching of doctors with professional actors in medical improvisation sessions over Zoom [74] are adequate examples of computer literacy with which clinicians should be equipped. Therefore, it is necessary to redefine and contextualize the definition of computer literacy in the domain of AI use and actively educate clinicians on this topic in practice and in training altogether.

The findings of this study have various practical implications. With an explicit focus on the clinicians' AI-assisted decision making, we found that their sensemaking of the diagnostic logic

provided by AI and the system features may pose a challenge to the health care workforce. Our findings correspond with a recent report by the National Academy of Medicine [75] that highlighted "augmented intelligence," with an emphasis on the supporting role of AI in data synthesis, interpretation, and decision making for health care multi-stakeholders, such as clinicians, patients, and other related professionals. In preparing clinicians for such a change, the report suggested that their training should incorporate education programs on how to assess and use AI products and services appropriately for new and incumbent professionals alike [75].

Moreover, to prevent AI algorithms from generating a simplified decision plan, health care providers should be involved in the case scenarios of patient diagnosis to enhance the effectiveness of the AI algorithms [36]. What is largely neglected in the discussion is the need to close gaps between practice and clinical training for increasing AI understanding. The developers of AI need to consider the clinicians' current levels of exposure to technology in practice and in simulation and then design clinician-centered algorithms and interfaces. To achieve the goal of human-AI collaboration in health care [76], beyond involving clinicians in the development of clinical AI algorithms, these algorithms and their interfaces should be more human-centered. There must be assessments of diagnoses made by AI in both practice and training to reduce the gap between theory and practice in the use of AI by the health care workforce. This can be done by gathering behavioral big data from clinicians in various disciplines to help align AI algorithms with the unique, subjective patterns of reasoning that humans employ in clinical diagnosis.

Limitations

As with all research, this study is not without limitations, and its results should be interpreted with caution. First, we employed purposive sampling techniques for the target population. Since our target audience was individuals with experience in clinical diagnosis in various scenarios, we targeted and recruited FNP students as our sample. As such, the response rate was relatively low, and missing values were prevalent in the data. Future research could benefit from increasing the sample size to compare group differences in greater depth. Although our sample represents the target population of this study, sampling at the national level would be beneficial for generalizing the results of this study. Second, our AI variable was operationalized as binary (1: AI assistance; 0: otherwise). Future research may consider a survey construct with items that capture the rich characteristics of multimedia technology variables in the research models. Lastly, as our results were derived from the use of AI-enabled diagnostic technology, our results may not be generalizable to other types of AI or other clinical decision-making categories.

Conclusions

In keeping with the recent interest in and expectations for AI-assisted decision making in health care, as a first step, our research explored clinicians' sentiments toward AI assistance and their perceptions using sentiment analysis and a mixed methods design. Our results indicated that, while there are negative or nonexistent effects of AI assistance in enhancing clinical decision making, clinicians have positive sentiments toward AI assistance in the decision-making process, comparable with their encounters with actual patients. With this potential, we suggest that health care leaders, policy makers, and AI developers need to collect clinicians' behavior data and revisit the design and user interface of AI to make it more clinician-centered and collaborative.

Authors' Contributions

HH created and distributed the survey, analyzed the results, and drafted the entire manuscript. DG created and distributed the survey and co-authored the manuscript. Both the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Tian S, Yang W, Grange J, Wang P, Huang W, Ye Z. Smart healthcare: making medical care more intelligent. Global Health Journal 2019 Sep;3(3):62-65 p. 206 [FREE Full text] [doi: 10.1016/j.glohj.2019.07.001]
2. Phillips KA. The coming era of precision health. Health Affairs 2021 Feb 01;40(2):363-363. [doi: 10.1377/hlthaff.2020.02393]
3. McCarthy J. What is artificial intelligence? Stanford University. 1998. URL: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html [accessed 2021-12-10]
4. 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]
5. Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017 Apr;69S:S36-S40 [FREE Full text] [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
6. Kim J, Heo W. Artificial intelligence video interviewing for employment: perspectives from applicants, companies, developer and academicians. Information Technology & People 2021 Apr 01. [doi: 10.1108/itp-04-2019-0173]
7. ANSI/CTA-2089.1-2020: Definitions/Characteristics of Artificial Intelligence in Health Care. Americal National Standards Institute (ANSI). 2020. URL: https://webstore.ansi.org/standards/ansi/ansicta20892020 [accessed 2021-12-10]

8. Padmanabhan P. AI in healthcare: The tech is here, the users are not. CIO. 2021 May 21. URL: https://www.cio.com/article/3619551/ai-in-healthcare-the-tech-is-here-the-users-are-not.html [accessed 2021-12-10]
9. Davenport T, Bean R. Optum Focuses On AI To Improve Administrative Decisions. Forbes. 2020 Oct 9. URL: https://www.forbes.com/sites/tomdavenport/2020/10/09/optum-focuses-on-ai-to-improve-administrative-decisions/ [accessed 2021-12-10]
10. Smith M, Higgs J, Ellis E. Factors influencing clinical decision making. In: Higgs J, Jensen GM, Loftus S, Christensen N, editors. Clinical reasoning in the health professions. Amsterdam, Netherlands: Elsevier; 2008:89-100.
11. Eisenberg JM. Sociologic influences on decision-making by clinicians. Ann Intern Med 1979 Jun 01;90(6):957-964. [doi: 10.7326/0003-4819-90-6-957] [Medline: 443692]
12. Case Study: Ochsner Health System - AI-powered Early-warning System Saves Lives. Americal Hospital Association. 2018 Jun. URL: https://www.aha.org/system/files/2018-06/ochner-value-initiative-warning-system-case-study.pdf [accessed 2021-12-10]
13. Zhao H, Li G, Feng W. Research on Application of Artificial Intelligence in Medical Education. 2018 Presented at: International Conference on Engineering Simulation and Intelligent Control (ESAIC); August 10-11, 2018; Hunan, China URL: https://ieeexplore.ieee.org/document/8530428 [doi: 10.1109/esaic.2018.00085]
14. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
15. Wang X, Liang G, Zhang Y, Blanton H, Bessinger Z, Jacobs N. Inconsistent performance of deep learning models on mammogram classification. J Am Coll Radiol 2020 Jun;17(6):796-803. [doi: 10.1016/j.jacr.2020.01.006] [Medline: 32068005]
16. 3rd Annual Optum Survey on AI in Health Care. Optum. 2020. URL: https://www.optum.com/business/resources/ai-in-healthcare/2020-ai-survey.html [accessed 2021-12-10]
17. Chenbrolu K, Ressler D, Varia H. Smart use of artificial intelligence in health care: Seizing opportunities in patient care and business activities. Deloitte. 2020. URL: https://www2.deloitte.com/us/en/insights/industry/health-care/artificial-intelligence-in-health-care.html [accessed 2021-12-10]
18. Lavender J. Investment in AI for healthcare soars. KPMG. 2018 Nov. URL: https://home.kpmg/xx/en/home/insights/2018/11/investment-in-ai-for-healthcare-soars.html [accessed 2021-12-10]
19. Maddox TM, Rumsfeld JS, Payne PRO. Questions for artificial intelligence in health care. JAMA 2019 Jan 01;321(1):31-32. [doi: 10.1001/jama.2018.18932] [Medline: 30535130]
20. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019 Dec;132:103971. [doi: 10.1016/j.ijmedinf.2019.103971] [Medline: 31630063]
21. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2018 Mar 19;294(1-2):567-592. [doi: 10.1007/s10479-018-2818-y]
22. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]
23. Baker A, Perov Y, Middleton K, Baxter J, Mullarkey D, Sangar D, et al. A comparison of artificial intelligence and human doctors for the purpose of triage and diagnosis. Front Artif Intell 2020 Nov 30;3:543405 [FREE Full text] [doi: 10.3389/frai.2020.543405] [Medline: 33733203]
24. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020 Mar 25;368:m689 [FREE Full text] [doi: 10.1136/bmj.m689] [Medline: 32213531]
25. Stanfill MH, Marc DT. Health information management: implications of artificial intelligence on healthcare data and information management. Yearb Med Inform 2019 Aug;28(1):56-64 [FREE Full text] [doi: 10.1055/s-0039-1677913] [Medline: 31419816]
26. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. Journal of Consumer Research 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
27. Shneiderman B. International Journal of Human–Computer Interaction 2020 Mar 23;36(6):495-504. [doi: 10.1080/10447318.2020.1741118]
28. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
29. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. Korean J Radiol 2019 Mar;20(3):405-410 [FREE Full text] [doi: 10.3348/kjr.2019.0025] [Medline: 30799571]
30. Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. Acad Med 2000 Feb;75(2):182-190. [doi: 10.1097/00001888-200002000-00020] [Medline: 10693854]

31. Kassirer JP. Teaching clinical reasoning: case-based and coached. Acad Med 2010 Jul;85(7):1118-1124. [doi: 10.1097/acm.0b013e3181d5dd0d] [Medline: 20603909]
32. Young ME, Dory V, Lubarsky S, Thomas A. How different theories of clinical reasoning influence teaching and assessment. Academic Medicine 2018;93(9):1415. [doi: 10.1097/acm.0000000000002303]
33. Offredy M. The application of decision making concepts by nurse practitioners in general practice. J Adv Nurs 1998 Nov;28(5):988-1000 [FREE Full text] [doi: 10.1046/j.1365-2648.1998.00823.x] [Medline: 9840870]
34. Innocent P, John R. Computer aided fuzzy medical diagnosis. Information Sciences 2004 May;162(2):81-104 [FREE Full text] [doi: 10.1016/j.ins.2004.03.003]
35. Lynn LA. Artificial intelligence systems for complex decision-making in acute care medicine: a review. Patient Saf Surg 2019;13:6 [FREE Full text] [doi: 10.1186/s13037-019-0188-2] [Medline: 30733829]
36. Grover S, Sengupta S, Chakraborti T, Mishra AP, Kambhampati S. RADAR: automated task planning for proactive decision support. Human–Computer Interaction 2020 Mar 19;35(5-6):387-412. [doi: 10.1080/07370024.2020.1726751]
37. Topol E. Deep medicine: how artificial intelligence can make healthcare human again. London, England: Hachette UK; 2019.
38. Shinners L, Aggar C, Grace S, Smith S. Exploring healthcare professionals' understanding and experiences of artificial intelligence technology use in the delivery of healthcare: An integrative review. Health Informatics J 2020 Jun 30;26(2):1225-1236 [FREE Full text] [doi: 10.1177/1460458219874641] [Medline: 31566454]
39. Van De Ven AH, Johnson PE. Knowledge for theory and practice. AMR 2006 Oct;31(4):802-821. [doi: 10.5465/amr.2006.22527385]
40. Nagel DA, Penner JL. Conceptualizing telehealth in nursing practice: advancing a conceptual model to fill a virtual gap. J Holist Nurs 2016 Mar 09;34(1):91-104. [doi: 10.1177/0898010115580236] [Medline: 25858897]
41. Bryce DA, King NJ, Graebner CF, Myers JH. Evaluation of a diagnostic reasoning program (DxR): exploring student perceptions and addressing faculty concerns. JIME 1998 May 12;1998(1):1. [doi: 10.5334/1998-1]
42. Saunders C, Ives B. DxR case study. CAIS 2000;3:1. [doi: 10.17705/1cais.00305]
43. Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Quarterly 1995 Jun;19(2):213. [doi: 10.2307/249689]
44. Agarwal R, Karahanna E. Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Quarterly 2000 Dec;24(4):665. [doi: 10.2307/3250951]
45. Venkatesh V, Thong JY, Xu X. 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]
46. Compeau DR, Higgins CA. Computer self-efficacy: development of a measure and initial test. MIS Quarterly 1995 Jun;19(2):189. [doi: 10.2307/249688]
47. Hah H, Goldin D, Ha S. The association between willingness of frontline care providers' to adaptively use telehealth technology and virtual service performance in provider-to-provider communication: quantitative study. J Med Internet Res 2019 Aug 29;21(8):e15087 [FREE Full text] [doi: 10.2196/15087] [Medline: 31469078]
48. Creswell JW. A Concise Introduction to Mixed Methods Research. Thousand Oaks, CA: Sage Publications, Inc; 2014.
49. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: Sage Publications, Inc; 2017.
50. Castro FG, Kellison JG, Boyd SJ, Kopak A. A methodology for conducting integrative mixed methods research and data analyses. J Mix Methods Res 2010 Sep 20;4(4):342-360 [FREE Full text] [doi: 10.1177/1558689810382916] [Medline: 22167325]
51. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 17;349(6245):261-266. [doi: 10.1126/science.aaa8685] [Medline: 26185244]
52. Canonico M, De Russis L. A comparison and critique of natural language understanding tools. Cloud Computing. 2018. URL: https://www.semanticscholar.org/paper/A-Comparison-and-Critique-of-Natural-Language-Tools-Canonico-Russis/fdf8bbd6dff6201c1965f81842852cbba59fe79e [accessed 2021-12-10]
53. Braun D, Mendez AH, Matthes F, Langen M. Evaluating natural language understanding services for conversational question answering systems. 2017 Presented at: 18th Annual SIGdial Meeting on Discourse and Dialogue; August 2017; Saarbrücken, Germany URL: https://aclanthology.org/W17-5522/ [doi: 10.18653/v1/w17-5522]
54. High R. The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works. IBM Redbooks. 2012. URL: http://www.redbooks.ibm.com/redpapers/pdfs/redp4955.pdf [accessed 2021-12-10]
55. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences). Thousand Oaks, CA: Sage Publications, Inc; 2002.
56. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford, England: Oxford University Press; 2009.
57. Choudhury MD, Lee MK, Zhu H, Shamma DA. Introduction to this special issue on unifying human computer interaction and artificial intelligence. Human–Computer Interaction 2020 May 31;35(5-6):355-361. [doi: 10.1080/07370024.2020.1744146]

58. LeRouge C, Mantzana V, Wilson EV. Healthcare information systems research, revelations and visions. European Journal of Information Systems 2017 Dec 19;16(6):669-671. [doi: 10.1057/palgrave.ejis.3000712]
59. LeRouge CM, Hah H, Deckard GJ, Jiang H. Designing for the co-use of consumer health technology in self-management of adolescent overweight and obesity: mixed methods qualitative study. JMIR Mhealth Uhealth 2020 Jun 29;8(6):e18391 [FREE Full text] [doi: 10.2196/18391] [Medline: 32597788]
60. Xu W. Toward human-centered AI. interactions 2019 Jun 26;26(4):42-46. [doi: 10.1145/3328485]
61. Chancellor S, Baumer EPS, De Choudhury M. Who is the "Human" in Human-Centered Machine Learning. Proc. ACM Hum.-Comput. Interact 2019 Nov 07;3(CSCW):1-32. [doi: 10.1145/3359249]
62. Dong L, Li W, Ma D, Lv C, Chen C. The evaluation and comparison on different types of resident doctors in training through DxR Clinician system. J. Phys.: Conf. Ser 2020 Jun 01;1549(4):042076. [doi: 10.1088/1742-6596/1549/4/042076]
63. Park Y, Jackson G, Foreman M, Gruen D, Hu J, Das A. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020 Oct;3(3):326-331 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa033] [Medline: 33215066]
64. Faggella D. Everyday Examples of Artificial Intelligence and Machine Learning. Emerj. 2020 Apr 11. URL: https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/ [accessed 2021-12-10]
65. Forster M. Six ways of experiencing information literacy in nursing: the findings of a phenomenographic study. Nurse Educ Today 2015 Jan;35(1):195-200 [FREE Full text] [doi: 10.1016/j.nedt.2014.06.005] [Medline: 25008026]
66. Scott S. Health literacy education in baccalaureate nursing programs in the United States. Nurs Educ Perspect 2016;37(3):153-158. [doi: 10.1097/01.nep.0000000000000005] [Medline: 27405196]
67. Skiba DJ. Nursing Education 2.0: YouTube™. Nursing Education Perspectives 2007;28(2):100-102 [FREE Full text]
68. Thompson B, Skiba D. Informatics in the nursing curriculum: a national survey of nursing informatics requirements in nursing curricula. Nurs Educ Perspect 2008;29(5):312-317. [Medline: 18834062]
69. Lin T. A computer literacy scale for newly enrolled nursing college students: development and validation. J Nurs Res 2011 Dec;19(4):305-317. [doi: 10.1097/JNR.0b013e318236d03f] [Medline: 22089656]
70. Campbell CJ, McDowell DE. Computer literacy of nurses in a community hospital: where are we today? J Contin Educ Nurs 2011 Aug;42(8):365-370. [doi: 10.3928/00220124-20110215-01] [Medline: 21366161]
71. Hobbs R. Create to learn: Introduction to digital literacy. Hoboken, NJ: Wiley-Blackwell; 2017.
72. Hobbs R. Reconceptualizing media literacy for the digital age. In: Martin A, Madigan D, editors. Literacies for learning in the Digital Age. London, England: Facets Press; 2006:99-109.
73. Pappano L. Training the next generation of doctors in Uganda. Internal Medicine News 2010 Oct;43(16):58-59. [doi: 10.1016/s1097-8690(10)70858-1]
74. Landi H. Healthcare AI investment will shift to these 5 areas in the next 2 years: survey. Fierce Healthcare. 2021 Mar 09. URL: https://www.fiercehealthcare.com/tech/healthcare-executives-want-ai-adoption-to-ramp-up-here-s-5-areas-they-plan-to-focus-future [accessed 2021-12-10]
75. Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of Medicine. JAMA 2020 Feb 11;323(6):509-510. [doi: 10.1001/jama.2019.21579] [Medline: 31845963]
76. Lee J. Is artificial intelligence better than human clinicians in predicting patient outcomes? J Med Internet Res 2020 Aug 26;22(8):e19918 [FREE Full text] [doi: 10.2196/19918] [Medline: 32845249]

Abbreviations

AI: artificial intelligence
CTA: Consumer Technology Association
FNP: family nurse practitioner
HIT: health information technology
HLM: hierarchical linear modeling
IS: information systems
NLU: natural language understanding

Edited by G Eysenbach; submitted 11.09.21; peer-reviewed by A Maroli, B Chaudhry, N Doreswamy; comments to author 01.10.21; revised version received 26.10.21; accepted 16.11.21; published 16.12.21.

Please cite as:
Hah H, Goldin DS
How Clinicians Perceive Artificial Intelligence–Assisted Technologies in Diagnostic Decision Making: Mixed Methods Approach
J Med Internet Res 2021;23(12):e33540
URL: https://www.jmir.org/2021/12/e33540
doi:10.2196/33540
PMID:34924356

Original Paper

A Novel Deep Learning–Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study

Li-Hung Yao[1*], MSc; Ka-Chun Leung[1*], MSc; Chu-Lin Tsai[2*], MD, ScD; Chien-Hua Huang[2*], MD, PhD; Li-Chen Fu[1*], PhD

[1]Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

[2]Department of Emergency Medicine, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan

[*]all authors contributed equally

Corresponding Author:
Li-Chen Fu, PhD
Department of Computer Science and Information Engineering
National Taiwan University
CSIE Der Tian Hall
No. 1, Sec. 4, Roosevelt Road
Taipei, 10617
Taiwan
Phone: 886 0935545846
Email: lichen@ntu.edu.tw

Abstract

Background: Emergency department (ED) crowding has resulted in delayed patient treatment and has become a universal health care problem. Although a triage system, such as the 5-level emergency severity index, somewhat improves the process of ED treatment, it still heavily relies on the nurse's subjective judgment and triages too many patients to emergency severity index level 3 in current practice. Hence, a system that can help clinicians accurately triage a patient's condition is imperative.

Objective: This study aims to develop a deep learning–based triage system using patients' ED electronic medical records to predict clinical outcomes after ED treatments.

Methods: We conducted a retrospective study using data from an open data set from the National Hospital Ambulatory Medical Care Survey from 2012 to 2016 and data from a local data set from the National Taiwan University Hospital from 2009 to 2015. In this study, we transformed structured data into text form and used convolutional neural networks combined with recurrent neural networks and attention mechanisms to accomplish the classification task. We evaluated our performance using area under the receiver operating characteristic curve (AUROC).

Results: A total of 118,602 patients from the National Hospital Ambulatory Medical Care Survey were included in this study for predicting hospitalization, and the accuracy and AUROC were 0.83 and 0.87, respectively. On the other hand, an external experiment was to use our own data set from the National Taiwan University Hospital that included 745,441 patients, where the accuracy and AUROC were similar, that is, 0.83 and 0.88, respectively. Moreover, to effectively evaluate the prediction quality of our proposed system, we also applied the model to other clinical outcomes, including mortality and admission to the intensive care unit, and the results showed that our proposed method was approximately 3% to 5% higher in accuracy than other conventional methods.

Conclusions: Our proposed method achieved better performance than the traditional method, and its implementation is relatively easy, it includes commonly used variables, and it is better suited for real-world clinical settings. It is our future work to validate our novel deep learning–based triage algorithm with prospective clinical trials, and we hope to use it to guide resource allocation in a busy ED once the validation succeeds.

(J Med Internet Res 2021;23(12):e27008) doi:10.2196/27008

KEYWORDS

emergency department; triage system; deep learning; hospital admission; data to text; electronic health record

Introduction

Background

Overcrowding in the emergency department (ED) is already a global public health issue and is clearly an important patient safety issue [1]. Many countries, such as Ireland, the United States, Canada, Germany, and Australia, have shown a continuous and significant increase in the number of ED visits [2-7]. In the United States, ED visits were estimated to increase from 136.9 million in 2015 to 145.6 million in 2016, an increase of 6.4%. The 10-year volume change was 24.7% and has increased by a total of 61.2% over the past 20 years (ED visits in 1996 were estimated at 90.3 million) [8,9]. In Taiwan, ED visits were estimated to increase from 7.18 million in 2017 to 7.64 million in 2019, an increase of 6.4%. In retrospect, the number of ED visits has increased by a total of 23.6% over the past 19 years [10].

The increasing number of ED visits has also caused a periodic imbalance in the supply and demand of ED and hospital resources, which leads to longer waiting times and delays in critical medical treatments. ED crowding is related to several adverse clinical outcomes, including higher mortality and morbidity [11,12]. Therefore, it is most important to design a method to properly identify urgent patients' priorities in the ED [13].

Related Work

Several research studies have focused on developing a system for predicting hospital admissions based on the patient's ED electronic medical record (EMR) [14]. Among these studies, the National Hospital Ambulatory Medical Care Survey (NHAMCS) data set [9] is the most common data set to be analyzed. Despite using the NHAMCS data set, those studies might end up with different outcomes being achieved. Here, we briefly introduce some existing methods and implementation results, followed by a description of the concepts and methods that our system uses.

Gligorijevic et al [15] developed a system for predicting the number of resources that the patients would need. They built a bidirectional Long Short-Term Memory (biLSTM) model to extract continuous data features and medical text data features, which resulted in a binary model with prediction accuracy and area under the receiver operating characteristic curve (AUROC) of 0.792 and 0.879, respectively. Moreover, they showed that using nurses' notes can provide a significant improvement in the prediction accuracy in comparison with using only standard continuous and categorical data.

Zhang et al [16] constructed a method for analyzing the patients' reasons for a visit to predict hospital admission using principal component analysis and traditional natural language processing (NLP) combined with multilayer neural network models and logistic regression (LR) model. In their study, they tested the model using a 10-fold cross-validation method, and the AUROC was 0.84. Sun et al [17] used the chi-square test to select the association between hospital admission and various possible risk factors and inputted the extracted association features into LR model for training to develop a prediction model, which was used to predict whether a need for hospital admission exists for ED patients. The involved variables included demographics (age, sex, and ethnic group), ED visit or hospital admission in the preceding 3 months, arrival mode, patient acuity category of the ED visit, and coexisting chronic diseases (diabetes, hypertension, and dyslipidemia). The AUROC for their study was 0.85.

Graham et al [18] used 3 machine learning algorithms to create the following models: LR, gradient boosted machines, and decision trees; these models were validated using a 10-fold cross-validation method repeated 5 times, whereby the accuracy of the best result in the gradient boosted machines model was 0.8. It turns out that their study can help clinicians plan the allocation of resources in advance and avoid the bottleneck of patient congestion.

Wang et al [19] developed a data-driven and evidence-based triage method to quickly identify acute and severe patients and prevent the waste of limited resources because of overdiagnosis. They proposed an attention-based biLSTM called *DeepTriager*, which processes both structured data and textual data from a clinical record to predict an ED patient's acuity level. The method can not only predict the acuity and severity of the outpatient but can also provide visualizable and interpretable evidence on the clinical context to support decision-making. The AUROC for binary classification (acuity 1 and 2) can achieve 0.93, which is 0.03 higher than that of traditional machine learning methods.

Study Aim

The aim of this study is to establish an effective and efficient system for predicting whether patients will eventually require hospital admission to provide a reference to physicians to rank the priority of treatment of patients in advance. In this proposed system, our goal is to use both conventional structured data and unstructured data to design a binary classification model to help identify the hospitalization needs of the ED patient visits.

Methods

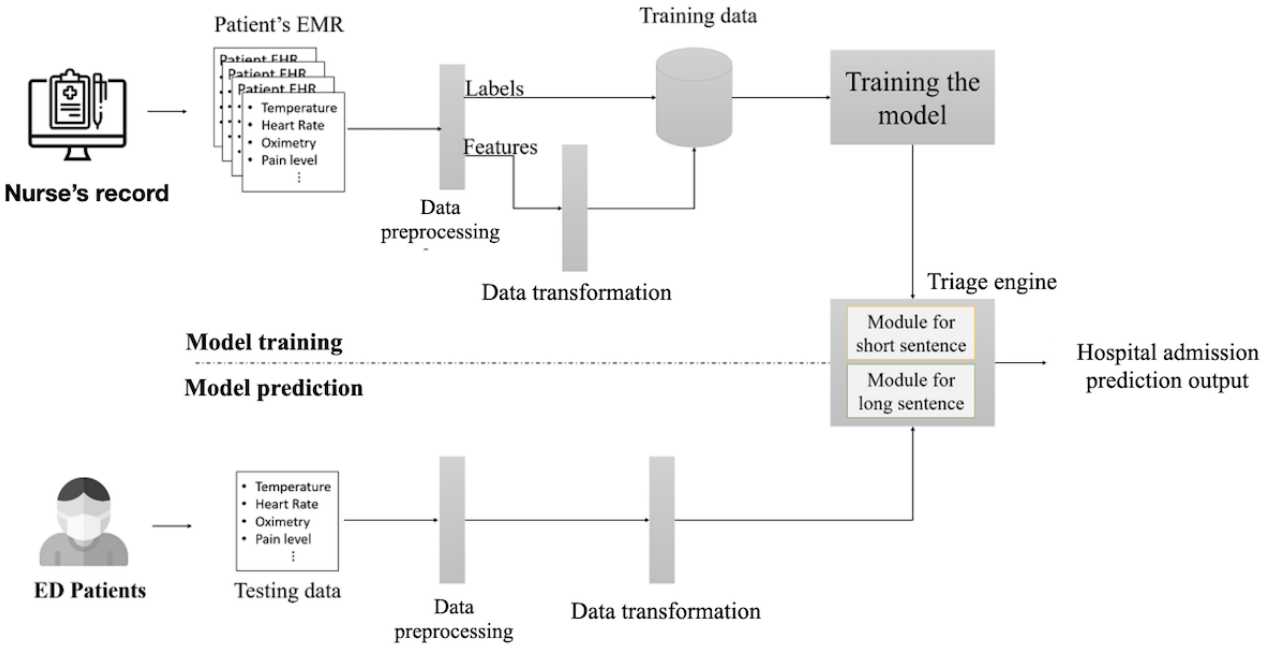
System Overview

This study focuses on establishing an effective and rapid system to predict whether patients will eventually need to be hospitalized to provide a reference to physicians to determine the priority of treatment of patients in advance. Moreover, to evaluate the effectiveness of our model for other clinical predictions, we also applied the model to other clinical outcomes and compared the obtained results with those from other algorithms.

The system overview in Figure 1 shows that our system is separated into two parts: the training part and the prediction part. The EMR values of each ED visit patient were used as input, and the patient's hospital admission decision from the physician was used as the ground truth. There were 3 steps in model training. The first step was to preprocess the input data, such as feature selection and filtering of the unusable or missing data, and the detailed method will be explained in the *Data Preparation* section. Next, the processed data were transformed into the corresponding text type. Finally, the transformed

transcript and ground truth were used to train the binary classification model. Once the training of the model was completed, it was tested against the unseen data. The unseen data were transformed to the text type, which was then fed into the model, and the output of the model was the probability of hospital admission for the ED patient visits.

Figure 1. System overview. EMR: electronic medical record; ED: emergency department.

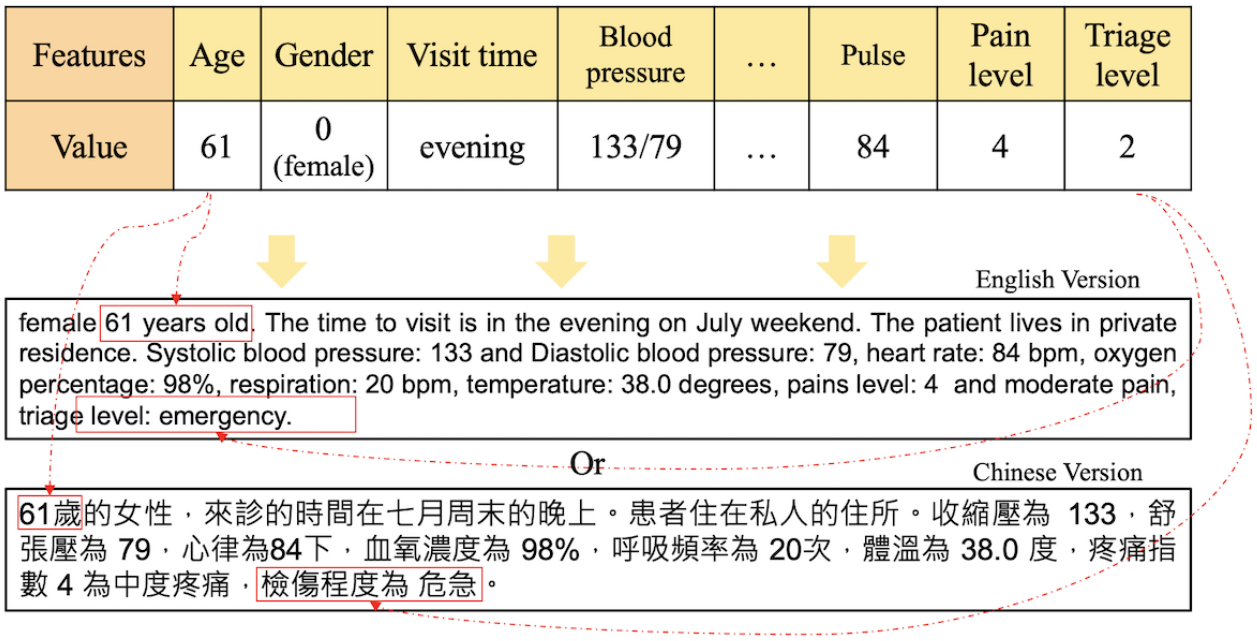


Data Transformation

To allow the mentioned data set to be more effectively handled by the proposed methods in our work, we first transformed the data into another text type. The method of transforming the original data into the text type is shown in Figure 2.

The table in Figure 2 shows the characteristics sampled from the original data set, including vital signs and other information, such as age, gender, blood pressure, oxygen, and pain index. The lower part of Figure 2 illustrates the format after data transformation in English or Chinese. Note that the original format of the data sample shown in Figure 2 reveals the features and their corresponding values. However, after transformation into text, all the feature names remained as words, but their corresponding values also appeared as words, so that the new format of the data sample now became a complete sentence. Then, we inputted the complete sentence into the model for training and analyzed the correlation between the features.

Figure 2. The method of data transformation.

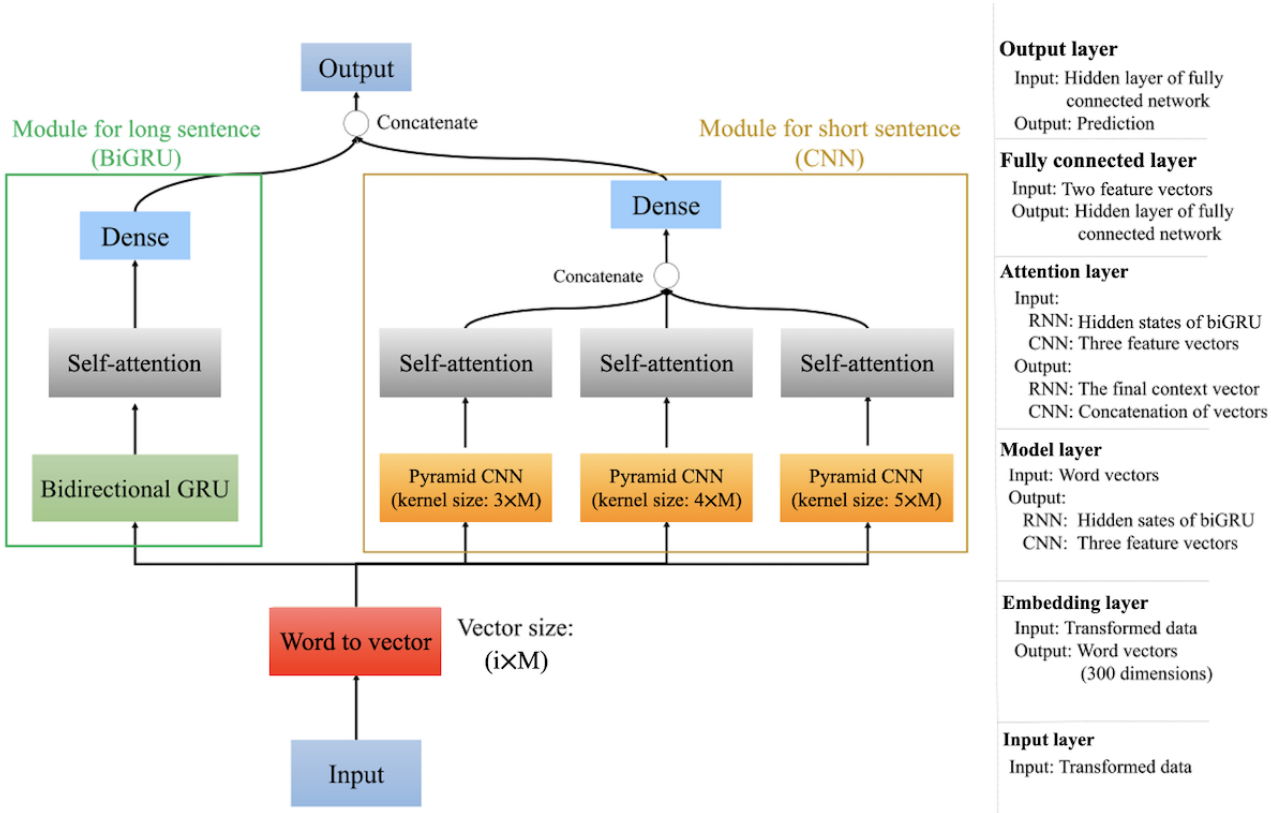


Triage Engine

Overview

In this work, our triage engine comprised 2 different parts, where the first part was used for analyzing short sentences (only convolutional neural network [CNN] type) and the second part was used for analyzing long sentences (only recurrent neural network [RNN] type). Experiments were conducted to verify the effectiveness of the triage engine after we first examined the performances of its 2 parts. Our classification engine was composed of 2 different text processing modules (ie, 2 parts). In the following section, we have introduced the characteristics of the 2 modules step by step and elaborated on the logic behind their design. The network architecture of our proposed triage engine, comprising the RNN-type module and the CNN-type module, is shown in Figure 3.

Figure 3. Network architecture of the triage engine. BiGRU: bidirectional gated recurrent unit; CNN: convolutional neural network; GRU: gated recurrent unit; RNN: recurrent neural network.



In our system, we used the data that had been converted into text format as input. As the text data contained many features, there was a certain relationship between the different features, such as the relationship between systolic blood pressure and diastolic blood pressure, pain index and pain location, and triage level and pain index. However, this relationship might have been lost after the features were converted into static word embeddings in high-dimensional space. Therefore, to be able to analyze the entire sentence for text information, we adopted the RNN structure, which has been shown to form a very useful algorithm. In fact, the RNN architecture is focused on the relationship among all the words from a sentence, and it is more appropriate to analyze the meaning of long sentences.

In addition, as the features and the corresponding values were converted into text sequentially in terms of a sentence, the name and value of a feature were in a neighboring relationship. To be able to accurately and effectively analyze the relationship among each feature and the corresponding value, the semantics of short sentences (among 3 words, 4 words, or 5 words) was very important, for example, *body temperature: 36.5 degrees*, *pain index: 5*, and *respiratory rate: 15 times*. In this short sentence analysis, CNN served as a very useful algorithm. This architecture focuses on the relationship between each word and its neighboring subjects. We will introduce it in detail later.

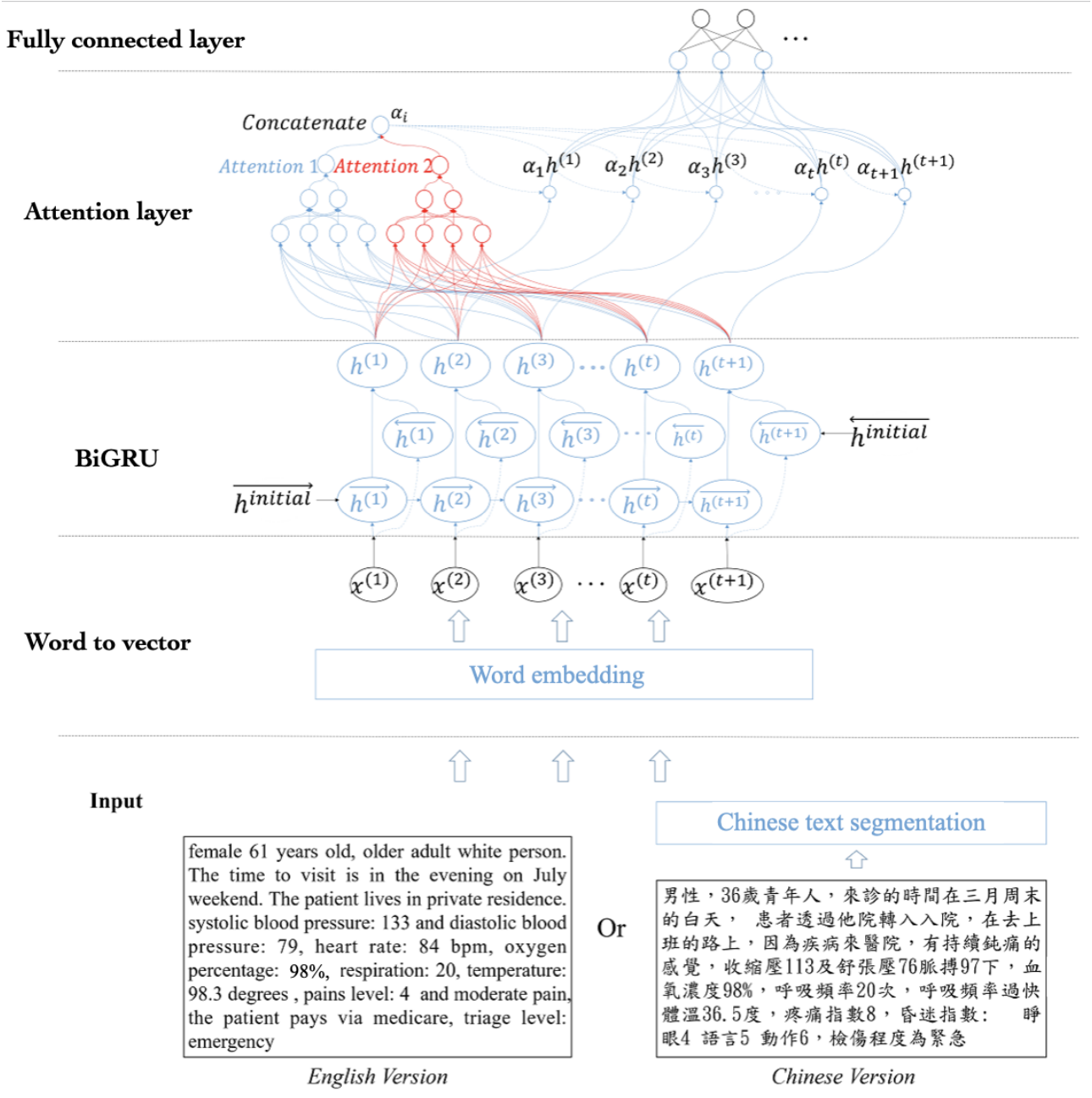
To effectively use the characteristics of the 2 learning algorithms of the 2 parts of the triage engine, we paid attention to the way in which the outputs of the 2 neural network models were effectively fused, which was also the focus of this system. In other words, we formed an overall model by merging the 2 parts to accomplish the single task, that is, prediction of the need for hospital admission. Our strategy was as follows: first, the 2 models were individually trained based on different data sets, and second, their parameters were optimized according to their respective losses and the corresponding parameter settings. After training the 2 models, we deleted the output layers of both the models and concatenated the last 2 fully connected layers, which were located before the output layers of the 2 models and the new output layer. Then, we fine-tuned the overall triage model on the 2 data sets, and the dropout was executed before the output layer. Finally, the output of the triage model predicted the probability of hospital admission for each ED visit.

Module for a Long Sentence (RNN Type)

The focus of this part of the study was on how to analyze the integrity of the text data, where the strength of feature extraction was extensive as the correlation among the patterns of the different samples was searched. Such a correlation not only involves time but also involves space. By learning from the sequence of sentences, the resulting model was able to effectively process each of the complete textual data and thus possessed a memory attribute.

The input of this model was the textual data transformed from the structural and unstructured data (EMR) of the ED visits, and the output was the vector that included the probability of hospital admission for the ED visits. Figure 4 shows the network architecture of the RNN part of the triage engine.

Figure 4. Recurrent neural network-type part of the triage engine. BiGRU: Bidirectional gated recurrent unit.



As depicted in Figure 4, first, all the words transformed from the EMR were input into the word vector layer, which was used to convert each word into the corresponding word vector. In our work, the pretrained word vector library was FastText [20,21], which is the most popular and useful library for learning word embeddings and text classification. First, as our work processed 2 data sets, one in English from NHAMCS and another in Chinese from the National Taiwan University Hospital (NTUH), we decided to develop our prediction model in both English and Chinese versions. Second, the word vectors were sequentially passed to the bidirectional gated recurrent unit (biGRU), in which the hidden sequences in 2 directions were concatenated at each time stamp to form a new hidden sequence. In general, the biGRU can obtain the features of a text more effectively. Thus, the hidden states from the biGRU were fed to the attention layer to evaluate the weights of each

hidden state, and the dot product between the evaluated weight value and each hidden state was calculated. The attention layer was composed of 2 fully connected feed-forward neural networks, using exponential linear units [22] as the activation function. For the output of the attention layer, Softmax, developed by Goodfellow et al [23], was chosen as the activation function.

The attention layer in this work was used to find the key information content units in different sentences from each ED visit's record and assess whether the patients will eventually need hospitalization. Moreover, the 2-layer attention network was used in this work as it is more effective than one with only a single layer in all sentences. Thus, the proposed system with multiple attention layers was more effective for the subsequent evaluation of the prediction performance throughout the experiment. Finally, the output from the attention layer was fed into the fully connected layers with 64 neurons.

Module for a Short Sentence (CNN Type)

This part of the study focused on extracting the local features of the text. By extracting the keywords of the document or sentence as features and training the classifier based on these features, it was possible to effectively analyze the more important and critical contexts of the sentences.

Similarly, Figure 5 shows the network architecture of the CNN part module of the triage engine. An image-like vector, whose format is $I \in$ $^{\times}$, was obtained by stacking the word vectors that are converted from the original text. More specifically, *h* and *w* denoted the height (number of words) and width (dimension of the word vector) of the image, respectively. In particular, Chinese words were processed by text segmentation first and then passed to the embedding layer. Then, the word vectors were used to perform convolution operations with 3 kernels of different sizes, which were 3, 4, and 5. Different kernels were used to find the relationships among short words, that is, various correlations between words.

For the convolutional operation, we adopted the concept from the deep pyramid CNNs [24], which is a low-complexity CNN architecture for text categorization that can efficiently represent long-range associations in a text. Instead of using the original CNN for text processing [25], we applied a simple network architecture to obtain better accuracy by increasing the network depth without significantly increasing the computational cost. Figure 6 shows the network architecture of the aforementioned pyramid CNN for text. Owing to the problem of degradation, the shortcut connections were expected to facilitate every few stacked layers to more easily fit a desired underlying mapping, and such thoughts of shortcut connections were the key concept for the pyramid CNN architecture. Therefore, according to the idea, the deep pyramid CNN under different kernel sizes was used in our work.

Figure 5. Convolutional neural network–type part of the triage engine. CNN: convolutional neural network.

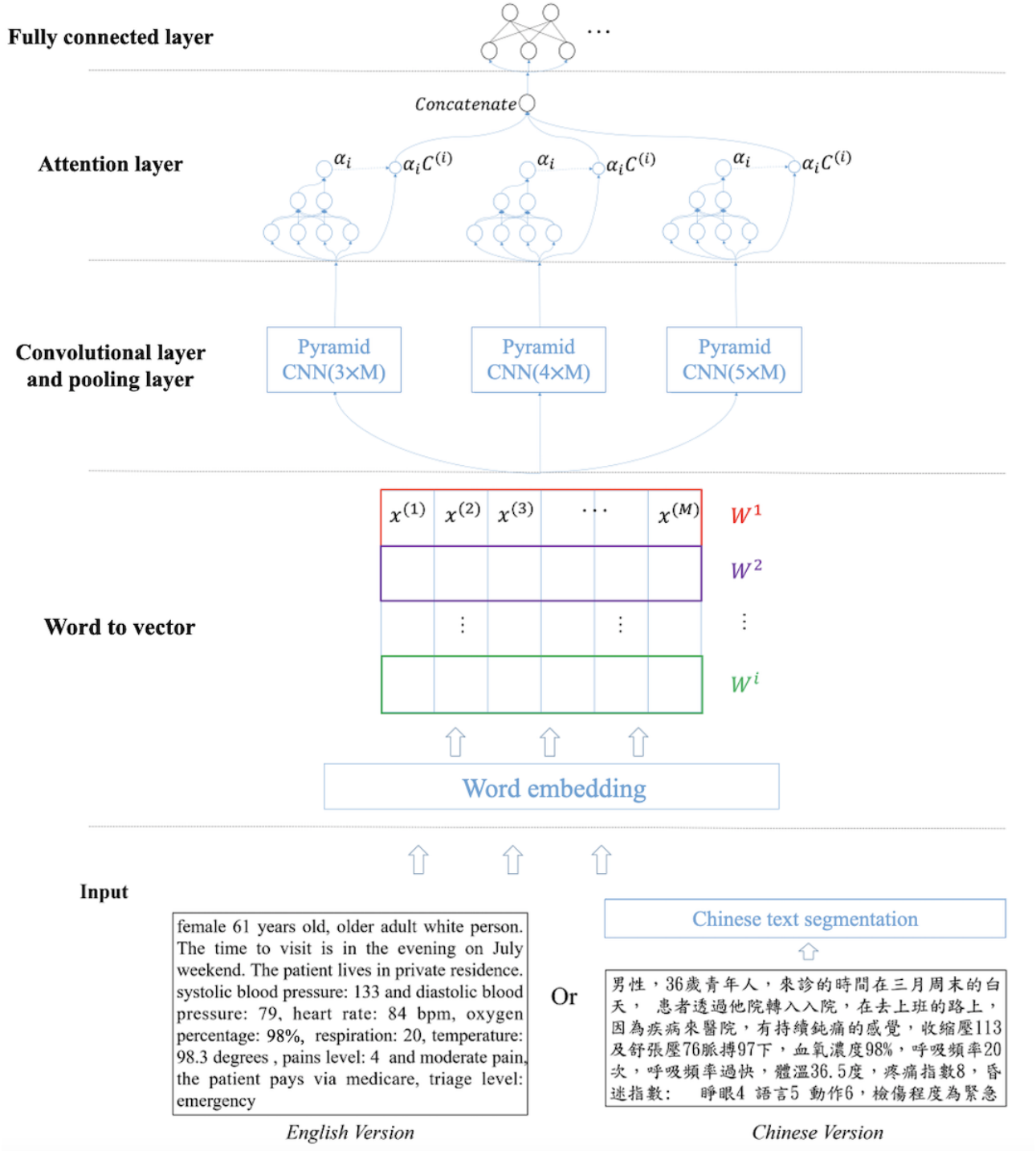
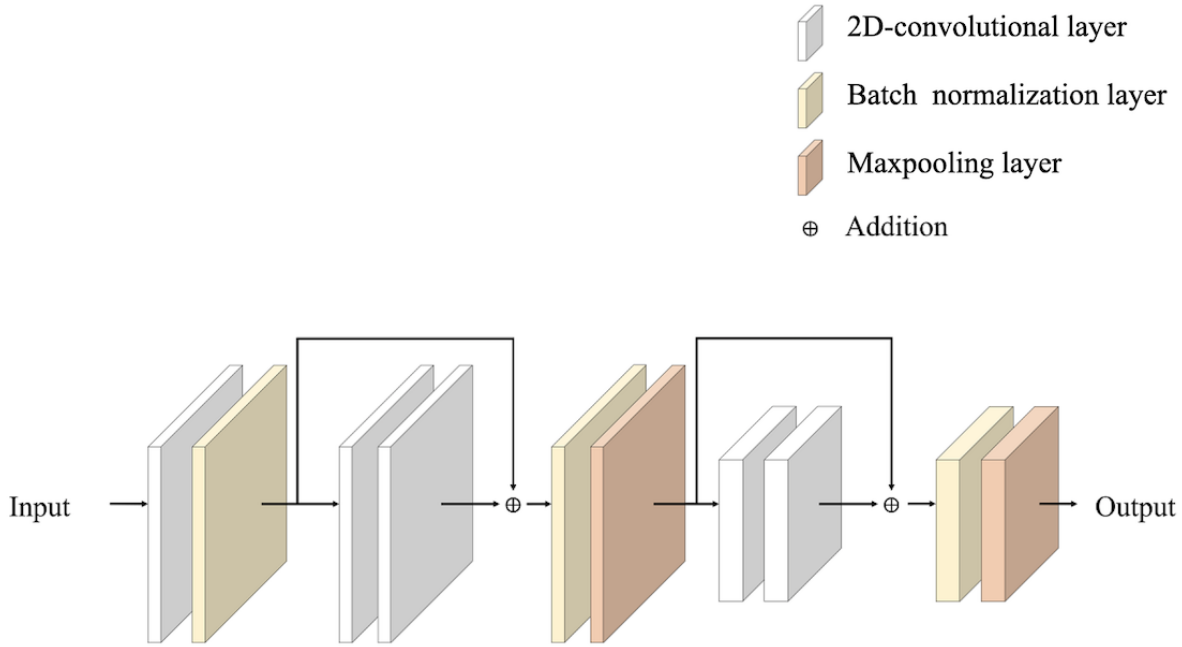


Figure 6. Architecture of pyramid convolutional neural network for text.



Model Synthesis by Integration of the 2 Modules

As shown in Figure 3, our final triage engine model was to integrate the 2 *pseudoengines* shown in Figures 4 and 5. Technically speaking, after the convolution operation, a feature map with the corresponding size was obtained, and each convolutional layer was followed by a max pooling layer of the corresponding size. The CNN-type part of the integrated engine was implemented to further distinguish the keywords from the transformed sentence. Then, the output of the max pooling layers was input to the self-attention layer, which was used to evaluate the weights of keyword vectors and calculate the dot product between the evaluated weight value and each keyword vector. Finally, the outputs from the 3 attention layers were concatenated and fed into fully connected layers with 64 neurons. The vector of the probability of hospital admission was calculated by applying the Softmax activation function in the fully connected layer, and we formulated the probability as follows:

$$Probability = CNN\ (k^3(X') + k^4(X') + k^5(X'))\ \textbf{(1)}$$

Model Training

The RNN-type module was trained with a learning rate of 0.00001 using an optimizer called Adam, developed by Kingma et al [26], which is a gradient descent method widely used in deep learning applications for computer vision and NLP. The batch size was set as 64, the number of iterations was set as 60, and the hidden states of the biGRU were set as 128. For the CNN-type module, the learning hyperparameters were the same as that of the RNN type. The size of the kernel was set to 3, 4, and 5, and the strides were set as 1. The loss function used for the integrated model was the cross-entropy sum between the predicted output and ground truth as follows:

$$l_{total} = l_{cnn} + l_{rnn}\ \textbf{(2)}$$

where y_i is the ground truth of class I, and $\hat{y}_i$ is the prediction of the model.

Results

Overview

A series of experiments were conducted to validate our design. To evaluate the effectiveness of our model, all the experiments were carefully conducted using stratified random sampling. The following procedure was performed separately on the NHAMCS and NTUH data sets. For internal comparison, 72% of the data were used as the training set, 8% of the data were used as the validation set, and 20% of the data were used as the hold-out testing set. The training and validation process was repeated 20 times, with 20 models generated, and the best-performing model in the hold-out testing set was selected as the final model. For external comparison, the same training and validation procedure was performed; however, only the best-performing model in the training and validation procedure was tested on the hold-out testing set to ensure a fair comparison.

Experiment Platform

We adopted Keras (Tensorflow-Graphics Processing Unit) to execute all the algorithms on computers with Nvidia GeForce GTX 1080Ti Graphics Processing Unit (with 11 GB RAM) and Intel Core i5 Central Processing Unit (with 64 GB RAM). In the processing of the loss function, we used the Adam optimizer with a learning rate of 0.00001, batch size of 64, and epochs of 60.

Data Preparation

In this study, 2 different data sets were used to evaluate the performance of the proposed system: the NHAMCS data set and the NTUH data set.

NHAMCS Data Set

In our study, the data from 118,602 ED patient visits collected between 2012 and 2016 were used. We selected 37 features, including *month*, *week*, *arrival time*, *age*, *residence*, *sex*, *race*, *did he or she come by ambulance*, *pay by insurance*, *pay by Medicare*, *pay by Medicaid*, *pay by work compensation*, *pay by self*, *no charge to pay*, *temperature*, *heart rate*, *respiratory rate*, *systolic blood pressure*, *diastolic blood pressure*, *pulse oximetry*, *pain scale*, *triage level*, *been ED during last 72 hours*, *dementia*, *cancer*, *cerebrovascular*, *COPD*, *heart failure*, *HIV*, *ECG*, *X-ray*, *CT-scan*, *MRI*, *Ultrasound*, *CPR*, *admitted to ICU*, and *hospital admission*.

NTUH Data Set

In our study, the data from 745,441 ED patient visits collected between 2013 and 2017 were used. We selected 31 features, including *age*, *sex*, *day zone*, *weekend*, *month*, *is he or she getting fever?*, *clinics by*, *clinics for*, *is job-related?*, *on the job way*, *pain character*, *pain period*, *CPR*, *ICU*, *acute change*, *account sequence number*, *systolic blood pressure*, *diastolic blood pressure*, *pulse*, *oxygen*, *respiration rate*, *body temperature*, *pain index*, *gcse*, *gcsv*, *gcsm*, *triage level*, *pain body part*, *pain period description*, *judgement description*, and *hospital admission*. All the features were recommended by Nottingham Trent University physicians.

Performance on the NHAMCS Data Set and Baseline

We verified our proposed fusion model using the NHAMCS data set and compared the results with the 2 parts of the model (RNNs and CNNs). As a result, the AUROC can achieve 0.872 using the proposed model. The other metrics of the performance of the proposed network are shown in Table 1. For our fusion model, the highest accuracy and specificity can reach 0.828 and 0.843, respectively.

So far, most existing studies have used different data sets. Here, to effectively evaluate the prediction quality of our model, we chose the traditional machine learning algorithms commonly used in other studies as our baselines for comparison, including LR, extreme gradient boosting (XGBoost), and random forest. Furthermore, we compared our model with the Bidirectional Encoder Representations From Transformers (BERT) [27] model, which is considered to be a milestone of NLP. Then, we compared the different results obtained from different methods under various metrics.

Table 2 shows 6 metrics of each algorithm. It can be seen that our proposed model scored the highest in 4 out of 6 metrics, including specificity, precision, accuracy, and AUROC, while comparing with other models. These results suggest that our proposed deep learning algorithm seems to be more promising than the traditional machine learning algorithms.

Table 1. Performance on the National Hospital Ambulatory Medical Care Survey data set using different methods.

Model	Sensitivity	Specificity	Accuracy	AUROC[a]
BiLSTM[b] only	0.756	0.768	0.767	0.850
BiLSTM+Att[c]	0.711	0.822	0.809	0.854
BiLSTM+2×Att	0.745	0.802	0.796	0.856
BiGRU[d] only	0.744	0.78	0.776	0.854
BiGRU+Att	0.757	0.804	0.798	0.863
BiGRU+2×Att	0.764	0.809	0.801	0.866
CNNs[e] (with 3 kernels)	0.756	0.768	0.767	0.85
Pyramid CNN (3 kernels)	0.727	0.813	0.804	0.855
Pyramid CNN (3 kernels) with attention layer	0.731	0.825	0.819	0.862
Our model	0.755	*0.843* [f]	*0.828*	*0.872*

[a]AUROC: area under the receiver operating characteristic curve.

[b]BiLSTM: bidirectional Long Short-Term Memory.

[c]Att: attention layer.

[d]BiGRU: bidirectional gated recurrent unit.

[e]CNN: convolutional neural network.

[f]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Table 2. Comparison with baseline algorithms in the National Hospital Ambulatory Medical Care Survey data set.

Model	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC[a]
Logistic regression	0.747	0.741	0.745	0.745	0.744	0.825
XGBoost[b]	0.761	0.736	0.749	0.748	0.748	0.834
Random forest	0.781	0.715	0.748	0.747	0.747	0.828
BERT[c]	0.789	0.768	0.773	0.781	0.779	0.852
Our model	0.755	*0.843* [d]	*0.818* [d]	0.759	*0.828*	*0.872*

[a]AUROC: area under the receiver operating characteristic curve.

[b]XGBoost: extreme gradient boosting.

[c]BERT: Bidirectional Encoder Representations From Transformers.

[d]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Performance on the NTUH Data Set and Baseline

We also verified our proposed fusion model using the NTUH data set and compared the results with the 2 parts of the model (RNNs and CNNs), which included 10 experiments. In the RNNs part, we experimented with 6 different combinations using biLSTM and biGRU with different layers of attention mechanisms to observe the changes in the 4 metrics (sensitivity, specificity, accuracy, and AUROC) under different combinations. In the CNN part, we experimented with 3 different combinations using a traditional CNN and pyramid CNN with an attention mechanism to observe the changes in the 4 metrics under different combinations. As a result, the AUROC can achieve 0.876 using the proposed model in our fusion model (Table 3).

Similarly, to effectively evaluate the prediction quality of our model, we chose 3 kinds of traditional machine learning algorithms commonly used as our baselines—LR, XGBoost, and random forest. Furthermore, we also compared our model with 2 common methods, deep neural network for structural data with biGRU for textual data and the BERT [27] model. Then, we compared the differences in the results between the different methods.

Table 4 shows 6 metrics of each algorithm. It can be seen that our proposed model outperforms all the other algorithms. The result of our proposed model suggested a great improvement in predicting hospitalization when compared with other traditional methods.

Table 3. Performance on the National Taiwan University Hospital data set using different methods.

Method	Sensitivity	Specificity	Accuracy	AUROC[a]
BiLSTM[b] only	0.748	0.792	0.77	0.848
BiLSTM+Att[c]	0.74	0.822	0.781	0.862
BiLSTM+2×Att	0.774	0.8	0.785	0.867
BiGRU[d] only	0.768	0.78	0.774	0.855
BiGRU+Att	0.805	0.767	0.786	0.866
BiGRU+2×Att	0.8	0.785	0.808	0.872
CNNs[e] (with 3 kernels)	0.78	0.803	0.791	0.868
Pyramid CNN (3 kernels)	0.784	0.793	0.798	0.868
Pyramid CNN (3 kernels) with attention layer	0.754	0.823	0.788	0.871
Our model	0.768	0.819	*0.825* [f]	*0.876*

[a]AUROC: area under the receiver operating characteristic curve.

[b]BiLSTM: bidirectional Long Short-Term Memory.

[c]ATT: attention layer.

[d]BiGRU: bidirectional gated recurrent unit.

[e]CNN: convolutional neural network.

[f]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Table 4. Comparison with baseline algorithms in the National Taiwan University Hospital data set.

Model	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC[a]
Logistic regression	0.705	0.805	0.758	0.755	0.756	0.83
XGBoost[b]	0.745	0.785	0.766	0.765	0.765	0.84
Random forest	0.739	0.784	0.762	0.761	0.762	0.84
DNN[c]+BiGRU[d]	0.744	0.775	0.771	0.766	0.771	0.858
BERT[e]	0.736	0.789	0.777	0.756	0.763	0.844
Our model	*0.768* [f]	*0.819*	*0.81*	*0.788*	*0.825*	*0.876*

[a]AUROC: area under the receiver operating characteristic curve.

[b]XGBoost: extreme gradient boosting.

[c]DNN: deep neural network.

[d]BiGRU: bidirectional gated recurrent unit.

[e]BERT: Bidirectional Encoder Representations From Transformers.

[f]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Discussion

Comparison With Other Related Studies

According to existing research on the prediction of hospitalization, most studies used specific feature selection methods combined with traditional machine learning algorithms. As shown in Table 5, their results show a variable performance on different metrics. In this section, to compare with other studies in a fair manner, only the best-performing model in the training and validation procedure was tested on the hold-out test set.

According to Table 5, our model achieved the highest performance in AUROC while being compared in the same open data set, that is, the NHAMCS data set. In addition, our work also achieved an excellent score in accuracy in comparison with private data sets.

Table 5. Performance of different research studies.

Study	Methods	Data set	Performance			
			Sensitivity	Specificity	Accuracy	AUROC[a]
Raita et al [28]	DNN[b]	NHAMCS[c]	0.79	0.71	—[d]	0.82
Zhang et al [16]	NLP[e]+PCA[f]+LR[g]	NHAMCS	—	—	—	0.846
Yan Sun et al [17]	LR	Private	—	—	—	0.849
Graham et al [18]	GBM[h]	Private	0.535	0.899	0.8	0.859
Our model	BiGRU[i]+ Att[j]+ PyCNN[k]	NHAMCS	0.654	*0.856* [l]	*0.834*	*0.856*
Our model	BiGRU+ Att+ PyCNN	NTUH[m]	0.606	0.852	0.806	*0.821*

[a]AUROC: area under the receiver operating characteristic curve.

[b]DNN: deep neural network.

[c]NHAMCS: National Hospital Ambulatory Medical Care Survey.

[d]Not available.

[e]NLP: natural language processing.

[f]PCA: principal component analysis.

[g]LR: logistic regression.

[h]GBM: gradient boosted machines.

[i]BiGRU: bidirectional gated recurrent unit.

[j]Att: attention layer.

[k]PyCNN: pyramid convolutional neural network.

[l]Italicization indicates that the best performance was shown by our model in the metric among the different models.

[m]NTUH: National Taiwan University Hospital.

Applying on Other Clinical Outcomes

To effectively evaluate the prediction quality of our model for other clinical results, we selected other common outcomes to test and set the results of traditional machine learning algorithms as our baselines and then compared the differences in the results between different methods in various metrics.

For Mortality Rate Prediction

As mortality rate has a high correlation with the emergency severity index (ESI) 5-level triage, we applied our model to predict the mortality rate on the NTUH data set, and the results are shown in Table 6. Owing to the small number of deceased patients, we chose not to test this data set because of convergence issues.

Compared with other algorithms, including the 3 traditional machine learning algorithms, our proposed model outperforms all other methods except in *sensitivity*.

Table 6. Performance of mortality rate prediction on the National Taiwan University Hospital data set.

Model	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC[a]
Logistic regression	0.903	0.887	0.895	0.895	0.896	0.954
XGBoost[b]	0.926	0.913	0.909	0.919	0.919	0.962
Random forest	0.933	0.898	0.915	0.915	0.916	0.958
Our model	0.917	*0.941* [c]	*0.939*	*0.928*	*0.941*	*0.983*

[a]AUROC: area under the receiver operating characteristic curve.

[b]XGBoost: extreme gradient boosting.

[c]Italicization indicates that the best performance was shown by our model in the metric among the different models.

For Prediction of Intensive Care Unit Admission

In the probability of intensive care unit admission, we tested our model on 2 data sets, and the results are shown in Tables 7 and 8. Table 7 shows the comparison of the 4 algorithms on the NHAMCS data set, and Table 8 shows the results of the 4 algorithms on the NTUH data set.

Similarly, compared with other algorithms, including the 3 traditional machine learning algorithms, our proposed model outperformed all other methods except in *sensitivity*.

Table 7. Performance of prediction of intensive care unit admission on the National Hospital Ambulatory Medical Care Survey data set.

Model	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC[a]
Logistic regression	0.787	0.734	0.761	0.760	0.761	0.845
XGBoost[b]	0.823	0.708	0.769	0.764	0.765	0.849
Random forest	0.876	0.707	0.800	0.790	0.792	0.861
Our model	0.805	*0.807* [c]	*0.807*	*0.806*	*0.824*	*0.884*

[a]AUROC: area under the receiver operating characteristic curve.

[b]XGBoost: extreme gradient boosting.

[c]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Table 8. Performance of prediction of intensive care unit admission on the National Taiwan University Hospital data set.

Model	Sensitivity	Specificity	Precision	F1 score	Accuracy	AUROC[a]
Logistic regression	0.811	0.846	0.829	0.828	0.828	0.905
XGBoost[b]	0.831	0.831	0.829	0.832	0.830	0.917
Random forest	0.833	0.828	0.831	0.830	0.831	0.911
Our model	0.823	*0.872* [c]	*0.865*	*0.843*	*0.870*	*0.920*

[a]AUROC: area under the receiver operating characteristic curve.

[b]XGBoost: extreme gradient boosting.

[c]Italicization indicates that the best performance was shown by our model in the metric among the different models.

Limitations

There are several limitations to this study. First, the NTUH database used in our experiment belongs to the NTUH and is not publicly available. Hence, it is hard to fairly compare it with the models developed in other studies. However, the NHAMCS data set is publicly available and may be used to evaluate the performance of the models across studies. Second, all the evaluations are based on retrospective data, and future prospective evaluation is needed.

Conclusions

ED crowding has become one of the biggest issues in health care services. Many countries have shown a steady but significant increase in the number of ED patient visits. Although the ESI system somewhat improves the process of treatment, it still relies on the nurse's judgment and is prone to the problem where most patients are triaged to ESI level 3. Moreover, the main purpose of the ESI is to classify patients and reserve the more limited resources for those belonging to the high-acuity classes who may need them more urgently. Therefore, a system that can help physicians accurately triage a patient's condition is imperative. In this work, we proposed a system based on the patients' ED EMR to predict the need for hospitalizations after the assigned procedures in the ED are completed. This system uses CNNs combined with RNNs, together with an attention mechanism for classification.

We validated the proposed triage engine based on the developed fusion model on 2 data sets, one of which is from an open data set (NHAMCS) that contains 118,602 ED patient visits in the United States, in which the accuracy and AUROC were 0.83 and 0.87, respectively. On the other hand, we also externally validated our work on the local NTUH data set that includes 745,441 ED patient visits in Taiwan, in which the accuracy and AUROC were 0.83 and 0.88, respectively. Moreover, to effectively evaluate the prediction ability of our proposed system, we also applied the model to other clinical outcomes, including mortality and admission to the intensive care unit. The results showed that our method is approximately 3% to 5% higher in accuracy than other common methods, including 3 traditional machine learning algorithms. Furthermore, the implementation of the proposed system is relatively easy, includes commonly used variables, and is better fitting for real-world clinical settings. It is our future work to validate our novel deep learning–based triage algorithm with prospective clinical trials, and we hope to use it to guide resource allocation in a busy ED once the validation succeeds.

The unstructured data used in this work were recorded manually by a nurse. However, the text information should be directly described by the ED visits during the ED clinical examination. Therefore, future work may focus on using automatic speech recognition to directly convert and use the speech data of the ED visits. Moreover, although our work includes an analysis of short and long sentences, it does not deal with the relevance of global words. Thus, our future works may focus on combining different types of deep learning algorithms in this system to provide a more comprehensive system, such as a graph convolutional network or transformer.

Acknowledgments

This research was supported by the Joint Research Center for AI Technology and All Vista Healthcare under the Ministry of Science and Technology of Taiwan under the grants 110-2634-F-002-042-, 110-2634-F-002-016-, 110-2634-F-002-046-, and 110-2634-F-002-049-, as well as by the Center for Artificial Intelligence and Advanced Robotics, National Taiwan University.

Conflicts of Interest

None declared.

References

1. Pines JM, Griffey RT. What we have learned from a decade of ED crowding research. Acad Emerg Med 2015 Aug;22(8):985-987 [FREE Full text] [doi: 10.1111/acem.12716] [Medline: 26194441]
2. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med 2013 Jun;61(6):605-611 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026] [Medline: 23218508]
3. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 2018;13(8):e0203316 [FREE Full text] [doi: 10.1371/journal.pone.0203316] [Medline: 30161242]
4. Morley C, Stankovich J, Peterson G, Kinsman L. Planning for the future: Emergency department presentation patterns in Tasmania, Australia. Int Emerg Nurs 2018 May;38:34-40 [FREE Full text] [doi: 10.1016/j.ienj.2017.09.001] [Medline: 28958418]
5. Guttmann A, Schull MJ, Vermeulen MJ, Stukel TA. Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. Br Med J 2011 Jun 01;342:d2983 [FREE Full text] [doi: 10.1136/bmj.d2983] [Medline: 21632665]
6. Dinh MM, Bein KJ, Latt M, Chalkley D, Muscatello D. Age before acuity: the drivers of demand for emergency department services in the Greater Sydney Area. Emerg Med J 2015 Sep;32(9):708-711. [doi: 10.1136/emermed-2014-204174] [Medline: 25532104]

7. Lowthian JA, Curtis AJ, Jolley DJ, Stoelwinder JU, McNeil JJ, Cameron PA. Demand at the emergency department front door: 10-year trends in presentations. Med J Aust 2012 Feb 06;196:128-132. [doi: 10.5694/mja11.10955] [Medline: 22304608]
8. Augustine J. The latest emergency department utilization numbers are in. ACEPNow. 2019. URL: https://www.acepnow.com/article/the-latest-emergency-department-utilization-numbers-are-in/ [accessed 2021-11-26]
9. Rui P, Kang K, Ashman J. National hospital ambulatory medical care survey: 2016 emergency department summary tables. Centers for Disease Control and Prevention. 2016. URL: https://www.cdc.gov/nchs/data/nhamcs/web_tables/2016_ed_web_tables.pdf [accessed 2021-11-27]
10. Statistics of Medical Care Institution's Status and Hospital Utilization 2019. Ministry of Health and Welfare, Taiwan. URL: https://www.mohw.gov.tw/lp-4932-2.html [accessed 2021-11-27]
11. Pines JM, Hilton JA, Weber EJ, Alkemade AJ, Al Shabanah H, Anderson PD, et al. International perspectives on emergency department crowding. Acad Emerg Med 2011 Dec;18(12):1358-1370 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01235.x] [Medline: 22168200]
12. Chalfin DB, Trzeciak S, Likourezos A, Baumann BM, Dellinger RP, DELAY-ED Study Group. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. Crit Care Med 2007 Jun;35(6):1477-1483. [doi: 10.1097/01.CCM.0000266585.74905.5A] [Medline: 17440421]
13. Ahsan KB, Alam MR, Morel DG, Karim MA. Emergency department resource optimisation for improved performance: a review. J Ind Eng Int 2019 Nov 27;15(S1):253-266. [doi: 10.1007/s40092-019-00335-x]
14. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
15. Gligorijevic D, Stojanovic J, Satz W, Stojkovic I, Schreyer K, Del Portal D, et al. Deep attention model for triage of emergency department patients. In: Proceedings of the 2018 SIAM International Conference on Data Mining (SDM). 2018 Presented at: SIAM International Conference on Data Mining (SDM); May 3-5, 2018; San deigo CA, USA. [doi: 10.1137/1.9781611975321.34]
16. Zhang X, Kim J, Patzer RE, Pitts SR, Patzer A, Schrager JD. Prediction of emergency department hospital admission based on natural language processing and neural networks. Methods Inf Med 2017 Oct 26;56(5):377-389. [doi: 10.3414/ME17-01-0024] [Medline: 28816338]
17. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. Acad Emerg Med 2011 Aug;18(8):844-850 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01125.x] [Medline: 21843220]
18. Graham B, Bond R, Quinn M, Mulvenna M. Using data mining to predict hospital admissions from the emergency department. IEEE Access 2018;6:10458-10469. [doi: 10.1109/access.2018.2808843]
19. Wang G, Liu X, Xie K, Chen N, Chen T. DeepTriager: a neural attention model for emergency triage with electronic health records. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2019 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov. 18-21, 2019; San Diego, CA, USA. [doi: 10.1109/bibm47256.2019.8983093]
20. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics; April 2017; Valencia, Spain p. 427-431. [doi: 10.18653/v1/e17-2068]
21. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Transact Assoc Computat Linguist 2017 Dec;5:135-146. [doi: 10.1162/tacl_a_00051]
22. Clevert D, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). In: Proceedings of the 4th International Conference on Learning Representations, ICLR. 2016 Presented at: 4th International Conference on Learning Representations, ICLR; May 2-4, 2016; San Juan, Puerto Rico p. 2-4 URL: https://arxiv.org/pdf/1511.07289.pdf%5cnhttp://arxiv.org/abs/1511.07289%5cnhttp://arxiv.org/abs/1511.07289.pdf
23. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, USA: The MIT Press; 2016.
24. Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics.: Association for Computational Linguistics; 2017 Presented at: 55th Annual Meeting of the Association for Computational Linguistics; July, 2017; Vancouver, Canada. [doi: 10.18653/v1/p17-1052]
25. Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1181]
26. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. 2015 Presented at: 3rd International Conference on Learning Representations; May 7-9, 2015; San Diego, CA URL: https://arxiv.org/abs/1412.6980

27. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June, 2019; Minneapolis, Minnesota p. 4171-4186. [doi: 10.18653/v1/n18-2]
28. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019 Feb 22;23(1):64 [FREE Full text] [doi: 10.1186/s13054-019-2351-7] [Medline: 30795786]

Abbreviations

AUROC: area under the receiver operating characteristic curve
BERT: Bidirectional Encoder Representations From Transformers
CNN: convolutional neural network
ED: emergency department
EMR: electronic medical record
ESI: emergency severity index
LR: logistic regression
NHAMCS: National Hospital Ambulatory Medical Care Survey
NLP: natural language processing
NTUH: National Taiwan University Hospital
RNN: recurrent neural network
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 07.01.21; peer-reviewed by A Teixeira, M Fernandes; comments to author 12.02.21; revised version received 21.03.21; accepted 26.10.21; published 27.12.21.

Please cite as:
Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC
A Novel Deep Learning–Based System for Triage in the Emergency Department Using Electronic Medical Records: Retrospective Cohort Study
J Med Internet Res 2021;23(12):e27008
URL: https://www.jmir.org/2021/12/e27008
doi:10.2196/27008
PMID:34958305

Original Paper

Recruitment of Patients With Amyotrophic Lateral Sclerosis for Clinical Trials and Epidemiological Studies: Descriptive Study of the National ALS Registry's Research Notification Mechanism

Paul Mehta[1], MD; Jaime Raymond[1], MPH; Moon Kwon Han[1], PhD, MPH, MSc; Theodore Larson[1], MSc; James D Berry[2], MD, MPH; Sabrina Paganoni[2,3], PhD, MD; Hiroshi Mitsumoto[4], MD; Richard Stanley Bedlack[5], PhD, MD; D Kevin Horton[1], DrPH, MSPH

[1]Agency for Toxic Substances and Disease Registry, Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Sean M Healey & AMG Center for ALS, Massachusetts General Hospital, Boston, MA, United States

[3]Spaulding Rehabilitation Hospital, Harvard Medical School, Boston, MA, United States

[4]Department of Neurology, Columbia University College of Physicians and Surgeons, New York City, NY, United States

[5]Department of Neurology, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Paul Mehta, MD
Agency for Toxic Substances and Disease Registry
Centers for Disease Control and Prevention
4770 Buford Hwy NE
Atlanta, GA, 30341
United States
Phone: 1 770 488 0556
Email: pum4@cdc.gov

Abstract

Background: Researchers face challenges in patient recruitment, especially for rare, fatal diseases such as amyotrophic lateral sclerosis (ALS). These challenges include obtaining sufficient statistical power as well as meeting eligibility requirements such as age, sex, and study proximity. Similarly, persons with ALS (PALS) face difficulty finding and enrolling in research studies for which they are eligible.

Objective: The aim of this study was to describe how the federal Agency for Toxic Substances and Disease Registry's (ATSDR) National ALS Registry is linking PALS to scientists who are conducting research, clinical trials, and epidemiological studies.

Methods: Through the Registry's online research notification mechanism (RNM), PALS can elect to be notified about new research opportunities. This mechanism allows researchers to upload a standardized application outlining their study design and objectives, and proof of Institutional Review Board approval. If the application is approved, ATSDR queries the Registry for PALS meeting the study's specific eligibility criteria, and then distributes the researcher's study material and contact information to PALS via email. PALS then need to contact the researcher directly to take part in any research. Such an approach allows ATSDR to protect the confidentiality of Registry enrollees.

Results: From 2013 to 2019, a total of 46 institutions around the United States and abroad have leveraged this tool and over 600,000 emails have been sent, resulting in over 2000 patients conservatively recruited for clinical trials and epidemiological studies. Patients between the ages of 60 and 69 had the highest level of participation, whereas those between the ages of 18 and 39 and aged over 80 had the lowest. More males participated (4170/7030, 59.32%) than females (2860/7030, 40.68%).

Conclusions: The National ALS Registry's RNM benefits PALS by connecting them to appropriate ALS research. Simultaneously, the system benefits researchers by expediting recruitment, increasing sample size, and efficiently identifying PALS meeting specific eligibility requirements. As more researchers learn about and use this mechanism, both PALS and researchers can hasten research and expand trial options for PALS.

(J Med Internet Res 2021;23(12):e28021) doi:10.2196/28021

KEYWORDS

amyotrophic lateral sclerosis; Lou Gehrig disease; motor neuron disease; clinical trials; patient recruitment; National ALS Registry; research notification mechanism

Introduction

Amyotrophic lateral sclerosis (ALS), commonly known as Lou Gehrig disease, is a progressive multifactorial neurodegenerative disease primarily affecting motor neurons. Conservative estimates suggest that approximately 17,000 Americans currently live with ALS, while 1500 new cases are diagnosed annually [1,2]. Most patients with ALS survive 2-5 years after receiving a diagnosis [3,4]. Although numerous treatments and therapeutic strategies are employed in the care of persons with ALS (PALS), there are only 2 approved medications that slow ALS progression: riluzole and edaravone. These drugs do not cure ALS, but rather modestly prolong survival or slow disease progression [5,6].

In 2008, the US Congress passed the ALS Registry Act which directed the federal Agency for Toxic Substances and Disease Registry (ATSDR) to create the National ALS Registry (Registry) [7]. The mission of the Registry is multifold and includes determining national epidemiological trends such as incidence, prevalence, and mortality; identifying and examining risk factors and potential etiologies; and facilitating and supporting ALS research [8].

Researchers face challenges in patient recruitment and enrollment, especially for rare, fatal diseases such as ALS [9]. These challenges include having sufficient sample size, satisfying narrow inclusion criteria (eg, disease duration <2 years, specific genetic mutations, forced vital capacity >70%), as well as meeting eligibility requirements such as age, sex, and study proximity [10-14]. Similarly PALS, like patients with other rare disorders, face difficulty finding and enrolling in research studies for which they are eligible [15]. In 2012, the Registry sought approval from the Centers for Disease Control and Prevention (CDC) Institutional Review Board (IRB) to use its database of patients as a recruitment tool for researchers [16]. The system is now called the research notification mechanism (RNM). The purpose of this system is to provide a tool for researchers to recruit for epidemiological, biomarker, and observational studies, as well as clinical trials. Our objective is to describe how ATSDR's National ALS Registry is connecting PALS with opportunities to participate in ALS research and trials.

Methods

Since its establishment in 2012, PALS enrolling in the ALS Registry have been provided the opportunity to consent to receive research notifications. Interested participants provide an email address for receipt of research notifications through the RNM. To use the RNM, ALS researchers seeking PALS for research or trial participation submit a completed application online at the National ALS Registry website [17]. In the application process, researchers can select specific eligibility criteria to screen patients. Such criteria could include age range, period since diagnosis, gender, family history of ALS, region (national, state, or city); however, some researchers may specify no screening criteria if their study could include all Registry participants (Table 1).

Table 1. Patient criteria under the National ALS Registry research notification application.

Prescreening	Criteria
Age range at diagnosis (years)	0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-75, >75
Diagnosis years	From (yyyy) to (yyyy)
Gender	Male or female
City (State) of US residence	eg, Atlanta, Georgia
Family history of ALS[a]	Mother, father, brother, sister, children
No prescreening needed	Materials sent to all participants taking part in the process

[a]ALS: amyotrophic lateral sclerosis.

Proof of the researcher's IRB approval is required; however, CDC IRB approval is not required as this system was previously approved. All applications are reviewed for completeness and researchers are contacted if items are outstanding (eg, missing curriculum vitae or application form). Applications are submitted to the Registry's review committee comprising of internal and external subject matter experts to determine scientific merit based on a set of established criteria (Table 2).

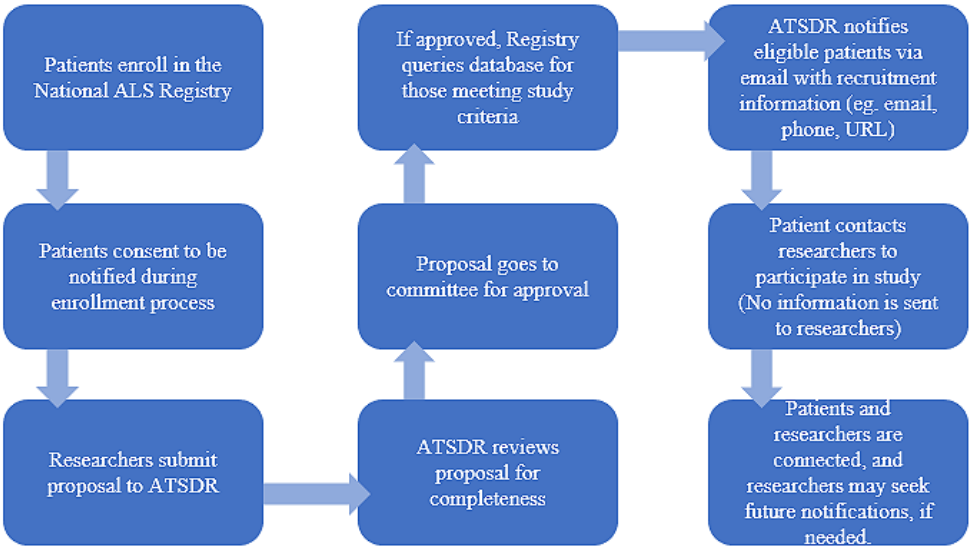
Table 2. Evaluation and approval of the application by the National ALS Registry review committee.

Evaluation	Criteria
1. Scientific merit (A-D is considered acceptable)	A. Outstanding/B. Excellent/C. Good/D. Acceptable/E. Unacceptable
2. Will the proposal provide useful information for ALS[a] patients? (A-C is considered acceptable)	A. High potential/B. Strong potential/C. Good potential/D. Limited potential/E. No apparent potential
3. Are the patient contact procedures and materials clear?	Yes/No/Requires more information
4. Are the patient contact procedures and materials appropriate, necessary, and sufficient?	Yes/No/Requires more information
5. In your judgment, would most ALS patients find the demands of the protocol reasonable?	Yes/No/Requires more information
6. Is there an acceptable risk/benefit ratio?	Yes/No/Requires more information
7. Are there adequate protections for patient confidentiality and privacy?	Yes/No/Requires more information
8. Does the proposal reach a satisfactory threshold for all 7 criteria listed above?	Yes/No
9. Do you support approval of the proposal as it has been submitted?	Yes/No
Comments for the investigators (limit of 500 words)	

[a]ALS: amyotrophic lateral sclerosis.

External committee members complete a conflict of interest form to ensure neutrality. Depending on the complexity of the application, approval time is within 3 weeks. If the committee cannot make a decision, final approval or disapproval resides with the ATSDR. Once the application is approved, Registry staff work with the researcher to coordinate a date and time for the notification, as shown in the process chart (Figure 1).

Figure 1. Flowchart of application approval process. ALS: amyotrophic lateral sclerosis; ATSDR: Agency for Toxic Substances and Disease Registry.



This communication helps the researcher to prepare for an influx of enrollment inquiries. The notification to PALS consists of a preapproved template with an attached document informing them of eligibility criteria (when applicable) and contact information for the researcher. A disclaimer is added to all notifications stipulating that CDC/ATSDR does not necessarily endorse the study and is not involved in its design or execution. PALS in receipt of the registration email can contact the research team directly to inquire about participation—no identifying information is shared with ALS researchers. RNM notifications can be sent to all registrants up to 3 times, if needed, and as requested by the researcher. In 2018, a fourth notification was implemented. The fourth communication follows the third round of notification to only newly registered patients after the 2018 implementation. All studies are posted on the Registry's website and are classified as either "Active" or "Closed." Contact information is also posted on the website for "Active" studies [18].

Results

Since the launch of the Registry in 2010, over 90% of enrollees have elected to be notified about new ALS clinical trials and epidemiological studies. From 2013 to 2019, 46 institutions,

both domestic and abroad, have leveraged this tool and 638,760 total emails were sent to consented patients. Annually, the greatest number of emails were sent in 2018 (n=293,422). In 2017, researchers at 10 institutions used the notification system—its broadest annual utilization. The median number of emails sent annually has been 41,433. The median number of emails per study was 8109 (Table 3).

Table 3. Number of emails sent to consented patients by Registry's RNM[a], 2013-2019.

Year	Studies (n=46), n (%)[b]	Emails (n=638,760), n (%)[c,d]	Yearly median
2013	4 (8.70)	5690 (0.89)	909
2014	8 (17.39)	27,035 (4.23)	5570
2015	9 (19.57)	41,433 (6.49)	5508
2016	3 (6.52)	20,007 (3.13)	2076
2017	10 (21.74)	82,114 (12.86)	9077
2018	6 (13.04)	293,422 (45.94)	30,091
2019	6 (13.04)	169,059 (26.47)	21,978

[a]RNM: research notification mechanism.

[b]Median number of studies for 2013-2019: 6.

[c]Median number of emails for 2013-2019: 41,433.

[d]Median number of emails per study: 8109.

Of the 7030 registrants who participated in the clinical notification, patients between the ages of 60 and 69 years had the highest level of participation (n=2417, 34.38%), whereas those between the ages of 18 and 39 and aged over 80 had the lowest (n=286, 4.07%, and n=218, 3.10%, respectively; Table 4). Age distribution in those who participated in clinical notification was found to be statistically different from the total registrants.

Table 4. Demographic characteristics of registrants who elected to receive notifications, January 1, 2013 to December 31, 2019.

Characteristic	Clinical notification participants (n=7030), n (%)	All registered participants (n=10,625), n (%)	*P* value[a]
Age at diagnosis, n (%)			<.001
18-39	286 (4.07)	341 (3.21)	
40-49	840 (11.95)	1041 (9.80)	
50-59	2067 (29.40)	2782 (26.18)	
60-69	2417 (34.38)	3817 (35.92)	
70-79	1202 (17.10)	2198 (20.69)	
80+	218 (3.10)	446 (4.20)	
Gender, n (%)			.71
Male	4170 (59.32)	6289 (59.19)	
Female	2860 (40.68)	4336 (40.81)	
Census region, n (%)			.18
Midwest	2076 (29.53)	3315 (31.20)	
South	2420 (34.42)	3480 (32.75)	
West	1464 (20.83)	2216 (20.86)	
East	1019 (14.50)	1556 (14.64)	
Other/Missing	51 (0.73)	58 (0.55)	

[a]*P* value, Cochran–Mantel–Haenszel test.

More males participated (4170/7030, 59.32%) than females (2860/7030, 40.68%). Furthermore, participation of 7030 patients elected to be notified for clinical and epidemiological studies varied by region of country: South at 34.42% (n=2420), Midwest at 29.53% (n=2076), West at 20.83% (n=1464), and the East at 14.50% (n=1019; Table 4). This representation is in parallel with what has been observed in all registered participants. Based on anecdotal feedback from researchers, it is conservatively estimated that over 2000 patients were recruited for clinical trials between 2013 and 2019. This was calculated using an estimated average of 50 patients recruited per study for 46 studies or 2300 patients recruited since 2013.

Some studies recruited more than 50 patients, whereas some less. Epidemiological studies conducted by Columbia University and University of Miami also demonstrated a higher percentage of patients recruited from the RNM than other sources (Tables 5 and 6).

Table 5. Utilization of the research notification mechanism by Columbia University[a].

Enrollment method	Recruited (n=227), n (%)	Enrolled (n=103), n (%)
Email Blast (ATSDR[b])	164 (72.2)	69 (66.9)
Pamphlet	21 (9.3)	13 (12.6)
Columbia University Irving Medical Center	35 (15.4)	20 (19.4)
ALSA[c] listserv	1 (0.4)	0 (0)
ALS[d] online forums	4 (1.8)	0 (0)
ATSDR conference	2 (0.9)	1 (1.0)

[a]Prospective comprehensive epidemiologic study in a large cohort in the National ALS Registry: identifying ALS risk factors.

[b]ATSDR: Agency for Toxic Substances and Disease Registry.

[c]ALSA: Amyotrophic Lateral Sclerosis Association.

[d]ALS: amyotrophic lateral sclerosis.

Table 6. Utilization of the research notification mechanism by the University of Miami[a].

Enrollment method	Mean enrollment rate per month	Unpaired *t* test[b] (*df*)	*P* value
Baseline	25		
Lecture	26	0.091 (17)	.92
Tweet	28	0.28 (22)	.78
Facebook	42	1.02 (20)	.32
National ALS[c] Registry Research Notification Tool	89	2.94 (30)	.009

[a]Rare Disease Clinical Research Network, Contact Registry for the Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) Consortium.

[b]Compared with baseline.

[c]ALS: amyotrophic lateral sclerosis.

A list of notable clinical trials and epidemiological studies is provided in Multimedia Appendix 1.

Since inception of the system, epidemiological studies have outnumbered clinical trials. These studies have focused on the evaluation of risk factors, genetics, and patient and caregiver burden, while clinical trials have focused on novel treatments to reverse or slow disease progression. A complete list of clinical trials and studies is available in the National ALS Registry website [18].

Discussion

Clinical trial recruitment for rare diseases such as ALS has evolved. Researchers are continuously looking for novel methods for outreach and recruitment. The launch of the Registry's RNM has helped to recruit patients for both clinical trials and epidemiological studies. The recruitment of participants for epidemiological studies and clinical trials is challenging [19,20]. This is especially the case for ALS where it is estimated that almost 60% of patients are not eligible for clinical trials [21]. The use of telemedicine to facilitate ALS clinical trials offers assistance with recruitment, consenting, and screening [22]. This is particularly evident with the COVID-19 pandemic and its impact on clinical research [23]. The cost of conducting clinical trials as well as the recruitment of patients are mitigating factors [24]. It is estimated that only 10% of the patient population with ALS participates in clinical trials [13].

Many factors contribute to the challenges of recruiting participants for ALS trials and a multifaceted approach is likely needed to improve research participation. Some models posit that behavior (trial enrollment) is determined by 3 main factors: motivation, access, and information [25]. Because ALS is a terminal disease, PALS are often highly motivated to enroll in trials. Unfortunately, access is still restricted for some ALS research. Almost 60% of PALS are deemed ineligible for most trials and access can be limited by mobility or geography [10,20]. The use of technology and the recent launch of the Healey ALS Platform Trial, which allows the testing of multiple treatments at once, should help reduce the cost of research, decrease trial time, and increase patient participation [26,27]. The Registry is poised to support this platform trial. This is evident by the growing list of clinical trials and epidemiological studies since 2013. The Registry's novel system provides a user-friendly mechanism for researchers to access an existing

pool of patients with ALS. Moreover, there is no charge to researchers to use the Registry to recruit patients with ALS.

Thousands of PALS have enrolled in research based on RNM notifications, and while the number of those who actually participate is challenging to determine with specificity, our estimate of 2300 PALS is based on conservative assumptions. In addition, the exact estimate of participants recruited by the Registry's system is limited because many studies have more than 1 route of recruitment and researchers do not typically ask where a patient is recruited from such as ClinicalTrials.gov or the National ALS Registry. There were more males among the notification participants than females, in line with the demographics of ALS, which affects more males than females [2,8,28]. With over 8000 notification emails sent per study, the RNM allows researchers to access a large pool of potential participants quickly and efficiently. Furthermore, sending multiple notifications (n=4) allows patients who may have missed the notification message in their email inbox to be reminded of opportunity to participate in clinical trials and epidemiological studies. Because emails have inherent limitations and to increase durability of the recruitment message, the Registry has also posted all active studies with respective contact information on the website. This gives researchers the maximum exposure for their study.

The Registry has recruited for several notable clinical trials. These include 2 Phase 2/3 clinical trials, 1 utilizing mesenchymal stem cell–neurotrophic factor cells and the other a coformulation of sodium phenylbutyrate–taurursodiol, that have shown promise as recently reported [29,30]. In addition, the Registry has provided support for several epidemiological studies. Studies from Columbia University and the University of Miami demonstrated that the notification system led enrollment efforts when compared with other methods such as print and social media.

There are limitations to the RNM; for example, the system does not recruit for all ALS clinical trials because investigators must submit an application for consideration. Besides, the RNM only recruits patients who have self-enrolled in the Registry and not those who have been identified via Centers for Medicare & Medicaid Services (CMS) or Veterans Affairs (VA) databases. The Registry is working with partner organizations such as the ALS Association, Muscular Dystrophy Association, and the Les Turner ALS Foundation to increase enrollment, especially for minority patients.

Patient registries have been found to be successful in the recruitment of clinical trials for cancers and other conditions globally [31]. The National ALS Registry's RNM is an effective tool for connecting PALS to researchers conducting clinical trials and epidemiological studies. The implementation and utilization of the RNM continues to benefit the researchers by helping to speed-up the recruitment process, increasing the study sample size, and easily and efficiently identifying patients meeting specific eligibility requirements. As more researchers learn about and use this mechanism, PALS and researchers can work together to accelerate therapy development for ALS.

Acknowledgments

The authors thank Dr Michael Benatar of the University of Miami for giving us the permission to include data on the CReATe Connect enrollment. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention, the Agency for Toxic Substances and Disease Registry, and the National Center for Environmental Health. The use of product names in this presentation does not constitute an endorsement of any manufacturer's product.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sampling of clinical trials and epidemiological studies utilizing the National ALS Registry's Research Notification Mechanism system, 2013-2019.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i12e28021_app1.xlsx]

References

1. Chiò A, Logroscino G, Traynor BJ, Collins J, Simeone JC, Goldstein LA, et al. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology 2013;41(2):118-130 [FREE Full text] [doi: 10.1159/000351153] [Medline: 23860588]
2. Mehta P, Kaye W, Raymond J, Punjani R, Larson T, Cohen J, et al. Prevalence of Amyotrophic Lateral Sclerosis - United States, 2015. MMWR Morb Mortal Wkly Rep 2018 Nov 23;67(46):1285-1289 [FREE Full text] [doi: 10.15585/mmwr.mm6746a1] [Medline: 30462626]
3. Mitsumoto H, Chad DA, Pioro EP. Amyotrophic Lateral Sclerosis. Philadelphia, PA: A F. A. Davis Company; Jan 02, 1998.
4. Eisen A. Amyotrophic lateral sclerosis is a multifactorial disease. Muscle Nerve 1995 Jul;18(7):741-752. [doi: 10.1002/mus.880180711] [Medline: 7783764]

5. Bhandari R, Kuhad A, Kuhad A. Edaravone: a new hope for deadly amyotrophic lateral sclerosis. Drugs Today (Barc) 2018 Jun;54(6):349-360. [doi: 10.1358/dot.2018.54.6.2828189] [Medline: 29998226]
6. Dharmadasa T, Kiernan MC. Riluzole, disease stage and survival in ALS. Lancet Neurol 2018 May;17(5):385-386 [FREE Full text] [doi: 10.1016/S1474-4422(18)30091-7] [Medline: 29525493]
7. ALS Registry Act Public Law. 2008. URL: https://www.cdc.gov/als/ALSRegistryActPublicLaw.html [accessed 2021-01-01]
8. Mehta P, Antao V, Kaye W, Sanchez M, Williamson D, Bryan L, et al. Prevalence of amyotrophic lateral sclerosis - United States, 2010-2011. MMWR Suppl 2014 Jul 25;63(7):1-14. [Medline: 25054277]
9. Atassi N, Yerramilli-Rao P, Szymonifka J, Yu H, Kearney M, Grasso D, et al. Analysis of start-up, retention, and adherence in ALS clinical trials. Neurology 2013 Sep 04;81(15):1350-1355. [doi: 10.1212/wnl.0b013e3182a823e0]
10. Petrov D, Mansfield C, Moussy A, Hermine O. ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment? Front Aging Neurosci 2017;9:68 [FREE Full text] [doi: 10.3389/fnagi.2017.00068] [Medline: 28382000]
11. ALS Association. VA Provides Benefits to Veterans with ALS and Families. 2020. URL: http://web.alsa.org/site/PageServer?pagename=va_provides_benefits.? [accessed 2021-01-01]
12. Horton DK, Graham S, Punjani R, Wilt G, Kaye W, Maginnis K, et al. A spatial analysis of amyotrophic lateral sclerosis (ALS) cases in the United States and their proximity to multidisciplinary ALS clinics, 2013. Amyotroph Lateral Scler Frontotemporal Degener 2018 Feb;19(1-2):126-133 [FREE Full text] [doi: 10.1080/21678421.2017.1406953] [Medline: 29262737]
13. Bedlack RS, Pastula D, Welsh E, Pulley D, Cudkowicz ME. Scrutinizing enrollment in ALS clinical trials: room for improvement? Amyotroph Lateral Scler 2008 Oct;9(5):257-265. [doi: 10.1080/17482960802195913] [Medline: 18608092]
14. Bedlack RS, Wicks P, Heywood J, Kasarskis E. Modifiable barriers to enrollment in American ALS research studies. Amyotroph Lateral Scler 2010 Dec;11(6):502-507. [doi: 10.3109/17482968.2010.484494] [Medline: 20446890]
15. Katyal N, Govindarajan R. Shortcomings in the Current Amyotrophic Lateral Sclerosis Trials and Potential Solutions for Improvement. Front Neurol 2017;8:521 [FREE Full text] [doi: 10.3389/fneur.2017.00521] [Medline: 29033893]
16. Malek AM, Stickler DE, Antao VC, Horton DK. The National ALS Registry: a recruitment tool for research. Muscle Nerve 2014 Nov;50(5):830-834 [FREE Full text] [doi: 10.1002/mus.24421] [Medline: 25111654]
17. Centers for Disease Control and Prevention. National ALS Registry. 2020. URL: https://www.cdc.gov/als/ [accessed 2021-01-01]
18. Centers for Disease Control and Prevention. ALS Research Notification for Clinical Trials and Studies. URL: https://www.cdc.gov/als/ALSResearchNotificationClinicalTrialsStudies_Active.html [accessed 2021-01-01]
19. Skiba MA, Bell RJ, Islam RM, Davis SR. Challenges in recruitment to an epidemiological study of young Australian women: the Grollo-Ruzzene Foundation Young Women's Health Study. Aust N Z J Public Health 2019 Apr;43(2):131-136. [doi: 10.1111/1753-6405.12868] [Medline: 30727023]
20. van Eijk RPA, Westeneng H, Nikolakopoulos S, Verhagen IE, van Es MA, Eijkemans MJC, et al. Refining eligibility criteria for amyotrophic lateral sclerosis clinical trials. Neurology 2019 Jan 09:e451-e460 [FREE Full text] [doi: 10.1212/WNL.0000000000006855] [Medline: 30626653]
21. New Eligibility Criteria Needed for ALS Clinical Trials. 2019 Jan 31. URL: https://www.neurologylive.com/view/new-eligibility-criteria-needed-for-als-clinical-trials [accessed 2021-01-01]
22. Govindarajan R, Berry JD, Paganoni S, Pulley MT, Simmons Z. Optimizing telemedicine to facilitate amyotrophic lateral sclerosis clinical trials. Muscle Nerve 2020 Sep;62(3):321-326. [doi: 10.1002/mus.26921] [Medline: 32415876]
23. Tuttle KR. Impact of the COVID-19 pandemic on clinical research. Nat Rev Nephrol 2020 Oct;16(10):562-564 [FREE Full text] [doi: 10.1038/s41581-020-00336-9] [Medline: 32760016]
24. Sullivan T. A Tough Road: Cost To Develop One New Drug Is $2.6 Billion; Approval Rate for Drugs Entering Clinical Development is Less Than 12%. 2019 Mar 21. URL: https://www.policymed.com/2014/12/a-tough-road-cost-to-develop-one-new-drug-is-26-billion-approval-rate-for-drugs-entering-clinical-de.html [accessed 2021-01-01]
25. Pateria D, Singh R. Patients Recruitment Forecast in Clinical Trials. Cognizant 20-20 Insights. Teaneck, NJ: Cognizant; 2015 Aug. URL: https://www.cognizant.com/whitepapers/patients-recruitment-forecast-in-clinical-trials-codex1382.pdf [accessed 2021-01-01]
26. Reynolds T. Clinical trials: can technology solve the problem of low recruitment? BMJ 2011 Jun 14;342:d3662. [doi: 10.1136/bmj.d3662] [Medline: 21672994]
27. Massachusetts General Hospital. Healey ALS Platform Trial. URL: https://www.massgeneral.org/neurology/als/research/platform-trial [accessed 2021-01-01]
28. Mehta P, Kaye W, Bryan L, Larson T, Copeland T, Wu J, et al. Prevalence of Amyotrophic Lateral Sclerosis - United States, 2012-2013. MMWR Surveill Summ 2016 Aug 05;65(8):1-12. [doi: 10.15585/mmwr.ss6508a1] [Medline: 27490513]
29. Berry JD, Cudkowicz ME, Windebank AJ, Staff NP, Owegi M, Nicholson K, et al. NurOwn, phase 2, randomized, clinical trial in patients with ALS. Neurology 2019 Nov 18;93(24):e2294-e2305. [doi: 10.1212/wnl.0000000000008620]

30. Paganoni S, Hendrix S, Dickson SP, Knowlton N, Macklin EA, Berry JD, et al. Long-term survival of participants in the CENTAUR trial of sodium phenylbutyrate-taurursodiol in amyotrophic lateral sclerosis. Muscle Nerve 2021 Jan;63(1):31-39 [FREE Full text] [doi: 10.1002/mus.27091] [Medline: 33063909]
31. Tan MH, Thomas M, MacEachern MP. Using registries to recruit subjects for clinical trials. Contemp Clin Trials 2015 Mar;41:31-38 [FREE Full text] [doi: 10.1016/j.cct.2014.12.012] [Medline: 25545027]

Abbreviations

ALS: amyotrophic lateral sclerosis
ATSDR: Agency for Toxic Substances and Disease Registry
CDC: Centers for Disease Control and Prevention
CMS: Centers for Medicare & Medicaid Services
CReATe: Clinical Research in ALS and Related Disorders for Therapeutic Development
IRB: Institutional Review Board
PALS: persons with amyotrophic lateral sclerosis
RNM: research notification mechanism
VA: Veterans Affairs

Edited by R Kukafka; submitted 17.02.21; peer-reviewed by C Franz, J Parkin Kullmann; comments to author 14.04.21; revised version received 21.04.21; accepted 22.09.21; published 07.12.21.

Please cite as:
Mehta P, Raymond J, Han MK, Larson T, Berry JD, Paganoni S, Mitsumoto H, Bedlack RS, Horton DK
Recruitment of Patients With Amyotrophic Lateral Sclerosis for Clinical Trials and Epidemiological Studies: Descriptive Study of the National ALS Registry's Research Notification Mechanism
J Med Internet Res 2021;23(12):e28021
URL: https://www.jmir.org/2021/12/e28021
doi:10.2196/28021
PMID:34878988

Original Paper

An Integrated, Scalable, Electronic Video Consent Process to Power Precision Health Research: Large, Population-Based, Cohort Implementation and Scalability Study

Clara Lajonchere[1*], PhD; Arash Naeim[2*], MD, PhD; Sarah Dry[3], MD; Neil Wenger[4], MPH, MD; David Elashoff[4], PhD; Sitaram Vangala[4], MSc; Antonia Petruse[5], BA; Maryam Ariannejad[1], BSc; Clara Magyar[3], PhD; Liliana Johansen[5], MPH; Gabriela Werre[5], MSHCD; Maxwell Kroloff[4], MD; Daniel Geschwind[1], MD, PhD

[1]Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[2]Center for SMART Health, Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[3]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[4]Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[5]Embedded Clinical Research and Innovation Unit, Clinical and Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Arash Naeim, MD, PhD
Center for SMART Health
Institute for Precision Health
David Geffen School of Medicine at UCLA
10911 Weyburn Ave
Suite 300e
Los Angeles, CA, 90095
United States
Phone: 1 3103670148
Email: anaeim@mednet.ucla.edu

Abstract

Background: Obtaining explicit consent from patients to use their remnant biological samples and deidentified clinical data for research is essential for advancing precision medicine.

Objective: We aimed to describe the operational implementation and scalability of an electronic universal consent process that was used to power an institutional precision health biobank across a large academic health system.

Methods: The University of California, Los Angeles, implemented the use of innovative electronic consent videos as the primary recruitment tool for precision health research. The consent videos targeted patients aged ≥18 years across ambulatory clinical laboratories, perioperative settings, and hospital settings. Each of these major areas had slightly different workflows and patient populations. Sociodemographic information, comorbidity data, health utilization data (ambulatory visits, emergency room visits, and hospital admissions), and consent decision data were collected.

Results: The consenting approach proved scalable across 22 clinical sites (hospital and ambulatory settings). Over 40,000 participants completed the consent process at a rate of 800 to 1000 patients per week over a 2-year time period. Participants were representative of the adult University of California, Los Angeles, Health population. The opt-in rates in the perioperative (16,500/22,519, 73.3%) and ambulatory clinics (2308/3390, 68.1%) were higher than those in clinical laboratories (7506/14,235, 52.7%; *P*<.001). Patients with higher medical acuity were more likely to opt in. The multivariate analyses showed that African American (odds ratio [OR] 0.53, 95% CI 0.49-0.58; *P*<.001), Asian (OR 0.72, 95% CI 0.68-0.77; *P*<.001), and multiple-race populations (OR 0.73, 95% CI 0.69-0.77; *P*<.001) were less likely to participate than White individuals.

Conclusions: This is one of the few large-scale, electronic video–based consent implementation programs that reports a 65.5% (26,314/40,144) average overall opt-in rate across a large academic health system. This rate is higher than those previously reported for email (3.6%) and electronic biobank (50%) informed consent rates. This study demonstrates a scalable recruitment approach for population health research.

(J Med Internet Res 2021;23(12):e31121) doi:10.2196/31121

KEYWORDS

biobanking; precision medicine; electronic consent; privacy; consent; patient privacy; clinical data; eHealth; recruitment; population health; data collection; research methods; video; research; validation; scalability

Introduction

Informed consent for the research use of data and biological specimens is an essential and critical component of a robust program in precision medicine [1-3]. Although the common rule [4] considers the research use of deidentified tissue as "not human subjects" research, the National Institutes of Health (NIH) Genomic Data Sharing Policy expects consent to be obtained for future research use and broad data sharing, even if biospecimens are deidentified [5,6].

The 2017 revision to the common rule includes a new category of regulatory broad consent that provides more flexibility for researchers to consent participants for the storage, biobanking, and secondary research use of identifiable information or biospecimens [4]. Further, many advocates and ethicists have articulated an obligation to communicate that remnant tissue may be used for research and that researchers should proactively obtain informed consent [7,8]. Patients have also expressed a desire to have their preferences dictate the use of clinical specimens for research [8,9]. From this perspective, large-scale precision medicine programs that hope to engender greater trust and foster external collaborations, including collaborations with commercial entities and federal agencies, should consider proactively structuring their consent processes to include these key aspects of a broader and more informative consent process.

The emergence of digital health has also played a significant role in defining precision health approaches to obtaining electronic consent [10-12]. Obtaining in-person paper consent is often resource intensive, is not easily scalable, and precludes digital responses from being incorporated in the electronic health record (EHR) and laboratory information management systems [10-14]. Given that precision medicine research requires large-scale patient engagement, innovations in consent processes and public education [1,15] are required. Animated video consent approaches have been effective in providing comprehensive information and improving participants' understanding of content [16,17] and have thus provided an opportunity to more effectively increase the participation of traditionally underrepresented communities, as content can be tailored to participant groups.

The Engaging University of California Stakeholders for Biorepository Research (EngageUC) Consent Trial [18] was conducted to harmonize biobanking policies and procedures across 5 University of California medical campuses with the help of NIH Clinical and Translational Science Institutes (CTSI). The investigators developed an evidence-based approach that allowed for iterative stakeholder engagement to develop efficient biobank operations and equitable governance processes. The key themes that emerged were that the public should be educated about biobanking; the source of consent content should be considered knowledgeable and trustworthy; the process should be low stress and provide an opportunity to obtain answers to questions; the format and language should be easy to understand; and stakeholders, including the community, should play a role in informing and advising the institution. This framework provides a pathway for addressing the technical and ethical challenges that must be resolved to ensure that biorepositories continue to support translational research in ways that are inclusive of the populations they serve.

In 2016, the University of California, Los Angeles (UCLA), Institute for Precision Health (IPH) launched their ATLAS Community Health Initiative to engage a diverse sample of patients across UCLA Health in precision health research. The goal was to create a powerful and robust clinical and genomic data resource for cutting-edge translational research. This required innovative and scalable solutions for sustaining the program's rapid growth. The IPH partnered with this study's team to further develop and pilot their electronic universal video consent (UCON) process for biological samples, which was used to power the ATLAS precision health biobank.

In the development phase of this program, we leveraged the learnings from the EngageUC Consent Trial [18] and engaged community members across the greater Los Angeles region and internal stakeholders across the UCLA Health system, David Geffen School of Medicine at UCLA, the IPH, and CTSI as part of the initial development and pilot of the UCON process in targeted clinics at UCLA Health [19]. This "one-time" consent process gives all adult patients the autonomy to choose whether they want their deidentified biospecimens and clinical data to be made available for research.

In this paper, we describe phase 2 of the ATLAS Community Health Initiative, which focuses on the operational implementation and scalability of the UCON process. This includes its interoperability with the EHR and laboratory information management systems that power the UCLA ATLAS precision health biobank. We expanded the animated UCON process to 18 UCLA Health clinics across the Los Angeles region to test its scalability as an enterprise solution.

Methods

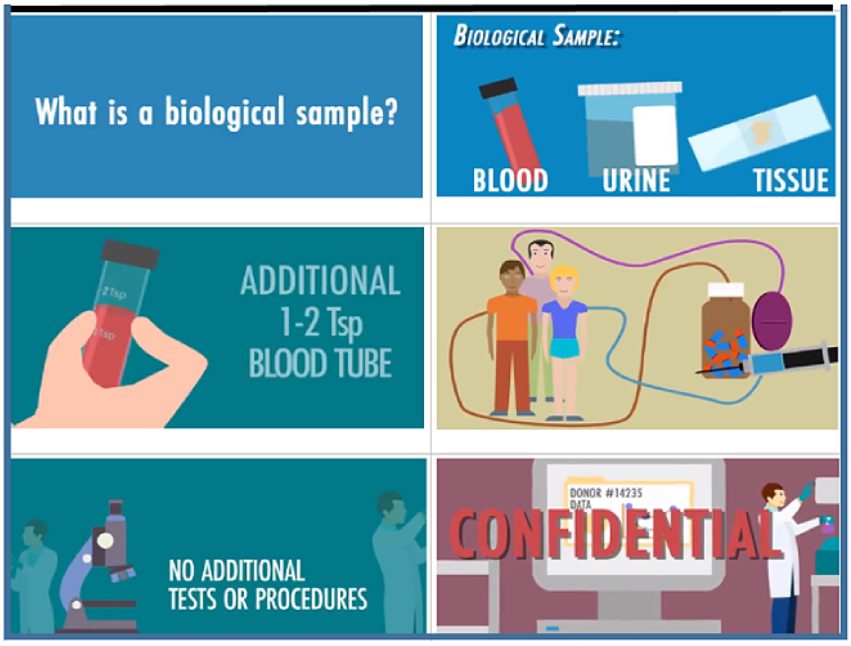
Study Setting

UCLA Health is one of the most diverse, comprehensive, and leading academic medical centers in Southern California. Ranked first in California and third in the nation, the UCLA Health system is comprised of a number of hospitals, including Ronald Reagan UCLA Medical Center and UCLA Santa Monica Medical Center, and an extensive primary care network of >180 medical offices in the greater Los Angeles area. This study was approved by the UCLA Institutional Review Board (institutional review board number: 15-001395IRB) with a waiver of written informed consent.

Electronic Video Consent

The electronic consent videos were designed to be self-administered, be fast (around 7 minutes), and meet the NIH consent threshold [19]. These fully animated (cartoon-like) videos were developed to better communicate content to lay audiences (Figure 1). The videos [20] were available in English and Spanish and included voice-overs. All of the essential components of regulatory broad informed consent were included in the videos. Our study team assembled a community advisory board (CAB) consisting of 11 respected leaders who were highly involved with organizations in the Los Angeles region that understood our diverse communities and represented their perspectives. The members were racially diverse (African American: n=2; Latinx: n=2; Asian American: n=1; Native American: n=1; Persian American: n=1; White and Non-Hispanic: n=4) and equitable with respect to gender (male: n=5; female: n=6). The CAB reviewed the results of a convenience sample of 117 patients who underwent cognitive testing. Operational feasibility was also tested with 625 additional patients. The CAB made the recommendation to move forward with the expansion to the broader UCLA Health population [19].

Figure 1. Representative screenshots of universal consent animated videos. Tsp: teaspoon.



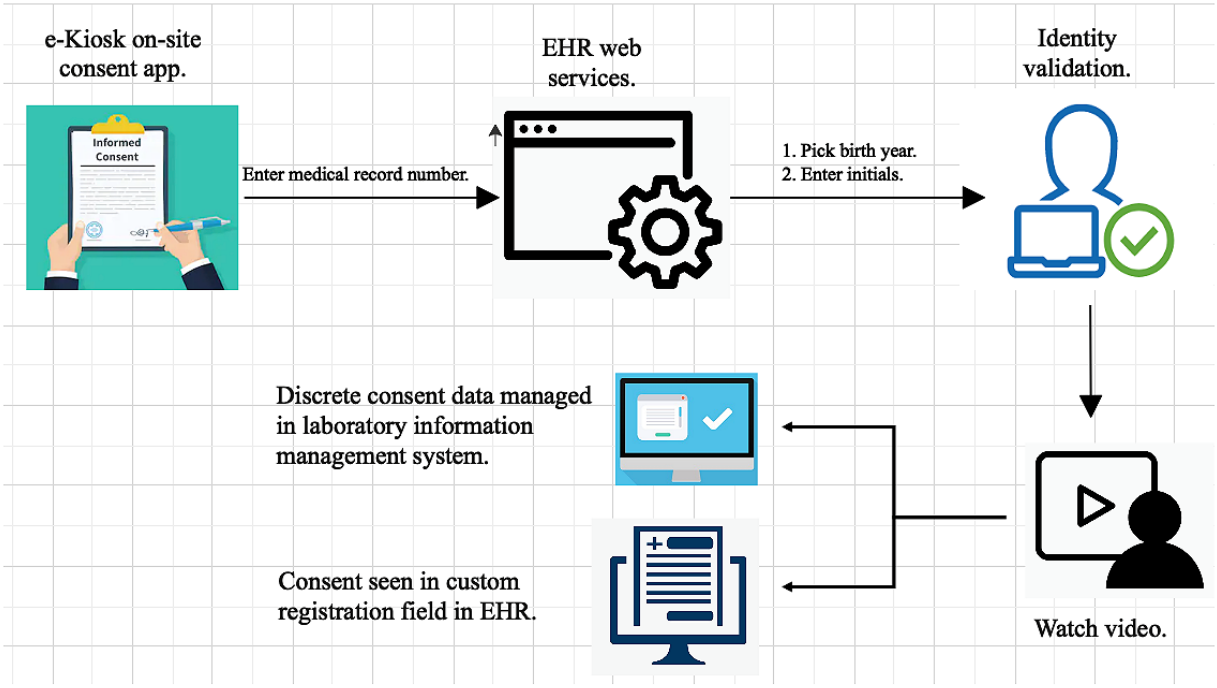
Consent Choices

With regard to patients' opt-in and opt-out statuses, patients could choose to (1) opt in to share their remnant samples plus a dedicated blood sample, (2) opt in to share remnant tissues only, or (3) opt out altogether. We also tracked the number of patients who opted in and indicated a willingness to be recontacted for future research.

Workflows

The UCON process was deployed in perioperative suites, clinical labs, and ambulatory clinics (Figure 2). In perioperative suites, patients were handed an iPad (Apple Inc), were instructed to watch the UCON video, and provided their consent decision. The iPad was then collected by nurses in the perioperative area. If a patient consented to providing an extra blood tube, the nurses placed the order in the EHR per the protocol. Patients typically waited for 30 to 60 minutes prior to their procedure, which provided ample time for completing the consent process. All extra tubes were automatically routed to the precision health biobank for DNA extraction and storage. In clinical labs, patients were handed an iPad by the front desk staff. Patients typically waited for 10 to 15 minutes prior to a lab draw and were able to watch the 5-minute UCON video and complete the consent process. Lastly, patients who were consented in ambulatory clinics completed the consent process at a self-service kiosk or by using an iPad before, during, or after a clinic visit. Patients waited for 10 to 15 minutes before moving to an exam room and another 10 to 15 minutes prior to meeting with a care provider, which provided ample time for completing the consent process. Once remnant (leftover) tissue samples became available for patients who opted in, these samples were accessioned into the ATLAS precision health biobank. DNA was extracted from whole blood and genotyped. Afterward, the remaining DNA was stored in the biobank.

Figure 2. Universal consent workflow. EHR: electronic health record.



Documentation and Tracking Within the EHR

After consent completion, consent decisions were transmitted to the precision health biobank's laboratory information management system (Biomaterial Tracking Management System; Daedalus Inc). PDF files containing patients' UCON completion statuses were sent to the EHR (Epic Systems Corporation) [21], which generated a receipt in real time in a custom registration field. The receipt ensured that patients would not be asked to complete the consent process a second time. Clinical staff used this field to determine which patients were eligible to complete the process.

Data Collection

The following data were collected from all patients who underwent the UCON process: sociodemographic information; comorbidity statuses based on the Charlson Comorbidity Index [22]; health utilization data, including ambulatory visits, emergency room visits, and hospital admissions; and consent decisions. With the exception of consent decisions, the data were all collected from the EHR by an honest broker who merged and deidentified the data via an institutional review board–approved process.

Statistical Analysis

Descriptive statistics were computed for all study variables. Quantitative variables were summarized by using means, SDs, and quartiles, while categorical variables were summarized by using frequencies and percentages. Sample characteristics were reported for the cohort that completed the consent process and for the larger UCLA Health patient population to determine whether there was sampling bias within and across clinics. US census data were used to obtain comparable summaries (where available) for residents of Los Angeles County.

Comparisons were made among the three workflows. Quantitative variables were compared by using 1-way analysis of variance *F* tests, and categorical variables were compared by using chi-square tests or Fisher exact tests, as appropriate. Consent status, which was categorized as *consented* or *declined*, was modeled via logistic regression.

A stepwise variable selection procedure was used to select a subset of predictor variables for inclusion in the model. This involved using entry and exit criteria for variables with a *P* value of <.001. The results of separate models that included 1 predictor at a time, as well as those of a multivariate model that combined all selected predictors, were reported. Odds ratios (ORs), 95% CIs, and *P* values were used to summarize model results. Model fit was evaluated by using the area under the curve. Statistical significance was defined as a *P* value of <.05. Cohort summaries and comparisons among cohorts were performed by using R version 3.5.0 (R Foundation for Statistical Computing). Logistic regression modeling was performed by using SAS version 9.4 (SAS Institute).

Results

Demographic Comparisons

The IPH consented 40,144 participants from March 2017 through June 2019 (Table 1). Of these participants, 26,314 (65.5%) opted into the ATLAS precision health biobank program. The demographics were representative of the UCLA Health population. ATLAS patients were only slightly older than the general UCLA Health population and much older than the Los Angeles County population; 30.2% (12,123/40,144) of ATLAS participants were aged ≥65 years, whereas 27.5% (160,479/583,282) and 16.6% of the UCLA Health and Los Angeles County populations, respectively, were aged ≥65 years.

The percentages of African Americans (2558/40,144, 6.4%) and Hispanic individuals (6055/40,144, 15.1%) in the ATLAS cohort were higher than those in the UCLA Health population (African Americans: 27,338/583,282, 4.7%; Hispanic individuals: 64,234/583,282, 11%) but lower than those in the Los Angeles County population (9.1% and 48.5%, respectively). The Los Angeles County data reflect census data from 2019, which only provides rates [23].

Table 1. Demographic and clinical characteristics of the cohort and reference populations.

Characteristics	Patients who completed the consent process (N=40,144)	UCLA[a] Health population (N=583,282)	Los Angeles County population[b]
Age (years), mean (SD)	53.1 (17.5)	51.6 (19.1)	—[c]
Aged ≥65 years, n (%)	12,123 (30.2)	160,479 (27.5)	16.6
Female, n (%)	22,302 (55.6)	333,998 (57.3)	50.7
Ethnicity, n (%)			
Not Hispanic or Latino	32,413 (80.7)	429,327 (73.6)	51.5
Hispanic or Latino	6055 (15.1)	64,234 (11)	48.5
Other, unknown, or refused	1676 (4.2)	89,721 (15.4)	0
Race, n (%)			
White	24,840 (61.9)	330,572 (56.7)	71
Asian	4667 (11.6)	54,043 (9.3)	15.1
Black or African American	2558 (6.4)	27,338 (4.7)	9.1
Native Hawaiian or other Pacific Islander	119 (0.3)	1204 (0.2)	0.4
American Indian or Alaska Native	115 (0.3)	2282 (0.4)	1.5
Multiple races, other, unknown, or refused	7845 (19.5)	167,843 (28.8)	3
Marital status, n (%)			
Married, significant other, or partner	22,897 (57)	289,364 (49.6)	—
Single	12,227 (30.5)	209,016 (35.8)	—
Divorced, separated, dissolved, or widowed	4573 (11.4)	52,201 (8.9)	—
Other or unknown	447 (1.1)	32,701 (5.6)	—
Neighborhood education level (percentage of high school graduates), n (%)			
<50%	17,074 (42.5)	209,697 (36)	—
>50%	17,785 (44.3)	256,249 (43.9)	—
Unknown	5285 (13.2)	117,336 (20.1)	—
Charlson score, mean (SD)	2.7 (3.5)	1.2 (2.3)	—
≥1 ambulatory visit, n (%)	39,870 (99.3)	547,283 (93.8)	—
≥1 inpatient admission, n (%)	12,438 (31)	43,695 (7.5)	—
≥1 emergency department–only visit, n (%)	9131 (22.7)	75,913 (13)	—
≥1 emergency department–to-inpatient visit, n (%)	5143 (12.8)	24,757 (4.2)	—

[a]UCLA: University of California, Los Angeles.

[b]The Los Angeles County data reflect census data from 2019, which only provides rates [23].

[c]Not available.

Finally, the IPH patients were more likely to be male, married, and well educated compared to the UCLA Health population (Table 1).

Although the UCON video was available in Spanish, most individuals of self-reported Hispanic descent elected to complete the consent process in English (39,294/40,144, 97.9%; Table 2).

Table 2. Consent outcomes by workflow.

Consent outcomes	Perioperative and admission workflows (n=22,519), n (%)	Lab workflow (n=14,235), n (%)	Other or unknown (n=3390), n (%)	*P* value
Consented in English	21,959 (97.5)	14,004 (98.4)	3331 (98.3)	<.001
Self-administered consent	21,672 (96.2)	13,989 (98.3)	3325 (98.1)	<.001
Consent status				<.001
Extra tube or saliva sample	10,640 (47.2)	4047 (28.4)	1361 (40.1)	
Remnant only	5860 (26)	3459 (24.3)	947 (27.9)	
Declined	6019 (26.7)	6729 (47.3)	1082 (31.9)	
Contact status				<.001
Do not contact	11,991 (53.2)	8166 (57.4)	1439 (42.4)	
Contact	9545 (42.4)	4716 (33.1)	1152 (34)	
Unknown	983 (4.4)	1353 (9.5)	799 (23.6)	

Differences in Consent Across Clinical Settings

We observed a marked difference in the willingness to opt in between patients presenting to different clinics (perioperative suites: 16,500/22,519, 73.3%; ambulatory setting: 2308/3390, 68.1%; Table 2) and patients in the laboratory medicine workflow (7506/14,235, 52.7%; *P*<.001). Patients in perioperative settings were more likely to share an extra tube (10,640/22,519, 47.2%) than patients in other clinic settings (*P*<.001). The number of patients who decided to opt in and provide a dedicated extra tube was highest in perioperative suites (10,640/22,519, 47.2%), second highest in ambulatory settings (1361/3390, 40.1%), and lowest in clinical labs (4047/14,235, 28.4%; *P*<.001). Further, the number of patients who were willing to be recontacted was highest in perioperative suites (9545/22,519, 42.4%), second highest in ambulatory settings (1152/3390, 34%), and lowest in clinical labs (4716/14,235, 33.1%).

Multivariate Analysis of Consent Variables

The most substantial differences in consent rates were observed in the perioperative setting. Those who were consented by using the laboratory workflow were significantly less likely to opt in than those who were consented by using the perioperative workflow (OR 0.43, 95% CI 0.42-0.46; *P*<.001; Table 3). There were also significant but less dramatic differences in opt-in rates based on self-reported ancestry; African American (OR 0.53, 95% CI 0.49-0.58; *P*<.001), Asian (OR 0.72, 95% CI 0.68-0.77; *P*<.001), and multiple-ethnicity populations (OR 0.73, 95% CI 0.69-0.77; *P*<.001) were less likely to opt in compared to those who self-identified as White (Table 3). More frequent health care utilization was also a significant predictor—those with greater than 1 inpatient admission had an OR of 1.28 (95% CI 1.22-1.35) for providing consent (*P*<.001). However, this was not the case for those who were frequent visitors of the emergency department, since those with greater than 1 emergency department visit had a lower consent rate than the average (OR 0.86, 95% CI 0.82-0.91; *P*<.001).

Table 3. Stepwise logistic regression model of any consent versus declines.

Variables	Unadjusted models		Multivariable model	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Race (reference: White)		<.001		<.001
Asian	0.66 (0.61-0.70)	<.001	0.72 (0.68-0.77)	<.001
Black or African American	0.52 (0.48-0.57)	<.001	0.53 (0.49-0.58)	<.001
Native Hawaiian or other Pacific Islander	1.06 (0.71-1.57)	.78	1.07 (0.71-1.60)	.76
American Indian or Alaska Native	0.77 (0.52-1.12)	.17	0.72 (0.49-1.07)	.10
Multiple races, other, unknown, or refused	0.67 (0.64-0.71)	<.001	0.73 (0.69-0.77)	<.001
Marital status (reference: married, significant other, or partner)		<.001		<.001
Single	0.81 (0.78-0.85)	<.001	0.89 (0.85-0.93)	<.001
Divorced, separated, dissolved, or widowed	1.00 (0.94-1.07)	>.99	0.95 (0.88-1.02)	.12
Other or unknown	0.60 (0.50-0.72)	<.001	1.00 (0.83-1.22)	.98
≥1 inpatient admission	1.41 (1.34-1.47)	<.001	1.28 (1.22-1.35)	<.001
≥1 emergency department–only visit	0.90 (0.86-0.94)	<.001	0.86 (0.82-0.91)	<.001
Workflow (reference: perioperative and admission workflows)		<.001		<.001
Lab workflow	0.41 (0.39-0.43)	<.001	0.43 (0.42-0.46)	<.001
Other or unknown	0.78 (0.72-0.84)	<.001	0.83 (0.76-0.89)	<.001
Consented in English	0.72 (0.62-0.84)	<.001	0.74 (0.63-0.87)	<.001

We also tracked the number of patients who reached out to the biobank with questions about their participation (n=13). The low number of patients may suggest that the UCON process was self-explanatory and did not require additional support to complete; however, this was not measured directly.

Discussion

In this study, we demonstrate that an integrated, institutional population–based electronic video consent process for the use of remnant biomaterials and deidentified phenotype data in research is feasible and scalable in large populations. The use of an animated, electronic video consent process is novel and innovative; ours is one of the few large-scale implementation studies that was conducted across a large health system and provided an enterprise solution for precision health research. Other reported efforts have been small, local, and nonrepresentative [24]. Compared to in-person paper consent, electronic video consent requires fewer human resources and less physical space and can be translated to reach diverse populations. A recent review of electronic consenting suggested that this modality is well received by participants, especially if it is accessible, user-friendly, and engaging and is tailored to specific patient populations [25,26]. The literature also suggests that this modality improves patient-centered outcomes, such as satisfaction and understanding [27]. At UCLA Health, we have implemented the translation of the UCON video into 8 languages that are represented widely across Los Angeles County. Other large precision health initiatives, such as the NIH All of Us Research Program, use more traditional consenting approaches and are limited to English and Spanish. In this regard, the expansion of the UCON process to several clinics across UCLA Health demonstrates the scalability of the UCON process as a recruitment engine that can power a large precision health program.

One of the goals of our precision health efforts is to ensure the inclusion of diverse populations and improve the accuracy of genetic prediction for all patients, regardless of ancestry. The inclusion of approximately 38.1% (15,304/40,144) of non-European patients in our biobank was likely achieved because we are located within one of the most diverse counties in the country. We note however that approximating the total diversity of Los Angeles County is a more daunting task that will require further, more significant outreach efforts. Although underrepresented populations are thought to have a lower willingness to participate in biobank research [28,29], community-based participatory research strategies have been shown to be effective [30-32]. Studies have shown that variations in the willingness to consent are mediated by different levels of trust in the health care and medical research system [33-35]. Patients are unsure about their rights over their biobank data [8] and have concerns about secondary research use [8,9]. Educating patients on the importance of genetic diversity for precision health approaches and creating toolkits to explain why their participation in large-scale programs is necessary will be important to the field.

This study also has its limitations. First, the UCON video was deemed to be at the ninth-grade reading level instead of the targeted seventh-grade reading level due to the use of many scientific terms and monosyllabic words. However, its success may reflect the demographics of the West Los Angeles population, which tends to be more educated and affluent. Despite this, we were successful in recruiting a diverse population for our expansion across UCLA Health. Future work

will include working with community partners at our affiliated community hospitals to adapt the UCON process to communities with lower health and linguistic literacy. Our data are also slightly skewed toward older patients, as they make up a large proportion of patients in our procedure units (ie, those in ambulatory, laboratory medicine, and perioperative settings). Although we identified some sociodemographic and health characteristics that predicted the willingness to opt into the ATLAS program, our multivariate models, while predictive, did not explain all of the variance. As such, it is likely that there are considerable unmeasured factors that influence an individual patient's decision process. The differences in consent rates across different clinical settings may suggest that patients are more willing to participate in biobank research in clinics, where there are higher rates of touch interactions between health care providers and patients.

Another potential limitation is that there was no measure for religiosity (how deeply religious an individual may be), which, in large national samples, has been shown to drive the willingness to participate in biobank research [36,37]. Additionally, our data only generalize to adults and do not generalize to pediatric patients. These populations often require specifically tailored consent and assent processes. Previously reported data show that patients' willingness to consent differs from their willingness to allow their child to participate in research [37]. However, adolescents have a high capacity for consent—similar to normal adults—and consider themselves capable of making voluntary choices [38]. Future work has been planned with pediatric patients within UCLA Health to shed light on this issue. Lastly, the consenting process has not been tested across multiple institutions but rather represents a single institutional study conducted at UCLA, which resides in the large metropolitan area of Los Angeles where approximately 79% of adults aged >25 years have a high school diploma [23]. Discussions are underway to expand the consenting process to other University of California sites where using a standard consent process is important for collaboration and data sharing (eg, when federal and state requirements change or when conducting international studies in which a high level of consent is required) [29,39-41].

In conclusion, the use of the UCON process is a scalable and highly efficient approach for population-based consenting and the recruitment of patients for precision health research. This study shows that the UCON process can be deployed to any number of devices and implemented at multiple medical locations, making it suitable for large-scale efforts with modest incremental costs. Implementation strategies using this approach need to balance obtaining a large sample of consenting participants with ensuring a representative sample. Moreover, since this consenting approach is largely self-administered, combining our consent process with community outreach and education may be essential to reaching underserved populations and those that may have strong health beliefs about participation in research activities. We believe that our consent video and process offer an approach that allows for the more robust inclusion of institutions that do not have the financial resources for using employees to obtain in-person consent. Given the current reality that many such institutions serve patients who are chronically ill, are of lower socioeconomic status, and are from underrepresented populations, our consent video and process offer the possibility for these groups to become better represented in precision medicine research.

Acknowledgments

The Clinical Translational Science Institute and Institute for Precision Health at the University of California, Los Angeles (UCLA), wish to thank all the patients who participated in this study. The authors would also like to acknowledge all of the members in the community advisory board. This study was supported by Federal Fund, National Institutes of Health, National Center for Advancing Translational Science (award number: UL1TR001881-01S2; grant title: *UCLA Clinical Translational Science Institute*).

Conflicts of Interest

None declared.

References

1. Ginsburg GS, Phillips KA. Precision medicine: From science to value. Health Aff (Millwood) 2018 May;37(5):694-701 [FREE Full text] [doi: 10.1377/hlthaff.2017.1624] [Medline: 29733705]
2. Petrini C. "Broad" consent, exceptions to consent and the question of using biological samples for research purposes different from the initial collection purpose. Soc Sci Med 2010 Jan;70(2):217-220. [doi: 10.1016/j.socscimed.2009.10.004] [Medline: 19853341]
3. Greely HT. Breaking the stalemate: a prospective regulatory framework for unforseen research uses of human tissue samples and health information. Wake Forest Law Rev 1999;34(3):737-766. [Medline: 12664910]
4. Federal policy for the protection of human subjects ('Common Rule'). U.S. Department of Health and Human Services, Office for Human Research Protections. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [accessed 2020-10-29]
5. NIH genomic data sharing policy. National Institutes of Health. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html [accessed 2020-10-29]

6. NIH policy supports broader sharing of genomic data, strengthens informed-consent rules: research participants must give consent for secondary sharing, even if data are de-identified. Am J Med Genet A 2015 Jan;167A(1):viii-viix. [doi: 10.1002/ajmg.a.36919] [Medline: 25528986]
7. Hens K, Nys H, Cassiman JJ, Dierickx K. The use of diagnostic collections of DNA for research: interviews at the eight Belgian centers for human genetics. Eur J Med Genet 2010;53(5):274-279. [doi: 10.1016/j.ejmg.2010.07.003] [Medline: 20624499]
8. De Vries RG, Tomlinson T, Kim HM, Krenz C, Haggerty D, Ryan KA, et al. Understanding the public's reservations about broad consent and study-by-study consent for donations to a biobank: Results of a national survey. PLoS One 2016 Jul 14;11(7):e0159113. [doi: 10.1371/journal.pone.0159113] [Medline: 27415017]
9. Vermeulen E, Schmidt MK, Aaronson NK, Kuenen M, van der Valk P, Sietses C, et al. Opt-out plus, the patients' choice: preferences of cancer patients concerning information and consent regimen for future research with biological samples archived in the context of treatment. J Clin Pathol 2009 Mar;62(3):275-278. [doi: 10.1136/jcp.2008.061069] [Medline: 19017681]
10. Rosa C, Campbell ANC, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54. [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
11. Simon CM, Schartz HA, Rosenthal GE, Eisenstein EL, Klein DW. Perspectives on electronic informed consent from patients underrepresented in research in the United States: A focus group study. J Empir Res Hum Res Ethics 2018 Oct;13(4):338-348. [doi: 10.1177/1556264618773883] [Medline: 29790410]
12. Boutin NT, Mathieu K, Hoffnagle AG, Allen NL, Castro VM, Morash M, et al. Implementation of electronic consent at a biobank: An opportunity for precision medicine research. J Pers Med 2016 Jun 09;6(2):17 [FREE Full text] [doi: 10.3390/jpm6020017] [Medline: 27294961]
13. Simon CM, Klein DW, Schartz HA. Interactive multimedia consent for biobanking: a randomized trial. Genet Med 2016 Jan;18(1):57-64 [FREE Full text] [doi: 10.1038/gim.2015.33] [Medline: 25834945]
14. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet Med 2016 Sep;18(9):906-913 [FREE Full text] [doi: 10.1038/gim.2015.187] [Medline: 26866580]
15. Weitzel KW, Alexander M, Bernhardt BA, Calman N, Carey DJ, Cavallari LH, IGNITE Network. The IGNITE network: a model for genomic medicine implementation and research. BMC Med Genomics 2016 Jan 05;9:1 [FREE Full text] [doi: 10.1186/s12920-015-0162-5] [Medline: 26729011]
16. Vanaken H. eConsent study provides insights to shape industry adoption. Appl Clin Trials 2016 Aug 01;25(8):12 [FREE Full text]
17. Abernethy ER, Campbell GP, Hianik R, Thomson MC, Dixon MD, Switchenko JM, et al. Videos improve patient understanding of misunderstood chemotherapy terms in a rural population. J Clin Oncol 2020 May 20;38(15_suppl):7022-7022. [doi: 10.1200/jco.2020.38.15_suppl.7022]
18. Dry SM, Garrett SB, Koenig BA, Brown AF, Burgess MM, Hult JR, et al. Community recommendations on biobank governance: Results from a deliberative community engagement in California. PLoS One 2017;12(2):e0172582. [doi: 10.1371/journal.pone.0172582] [Medline: 28235046]
19. Naeim A, Dry S, Elashoff D, Xie Z, Petruse A, Magyar C, et al. Electronic video consent to power precision research: A pilot cohort study. JMIR Form Res 2021 Sep 08;5(9):e29123 [FREE Full text] [doi: 10.2196/29123] [Medline: 34313247]
20. Universal consent: What is universal consent? Sharing Biological samples, precision medicine. UCLA Health. URL: https://www.uclahealth.org/precision-health/universal-consent [accessed 2021-11-03]
21. Software. Epic. URL: https://www.epic.com/software [accessed 2021-06-19]
22. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
23. U.S. Census Bureau QuickFacts: Los Angeles County, California. United States Census Bureau. 2019. URL: https://www.census.gov/quickfacts/losangelescountycalifornia [accessed 2021-11-03]
24. McDonald JA, Vadaparampil S, Bowen D, Magwood G, Obeid JS, Jefferson M, et al. Intentions to donate to a biobank in a national sample of African Americans. Public Health Genomics 2014;17(3):173-182 [FREE Full text] [doi: 10.1159/000360472] [Medline: 24942180]
25. Skelton E, Drey N, Rutherford M, Ayers S, Malamateniou C. Electronic consenting for conducting research remotely: A review of current practice and key recommendations for using e-consenting. Int J Med Inform 2020 Nov;143:104271 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104271] [Medline: 32979650]
26. Robins D, Brody R, Jeong IC, Parvanova I, Liu J, Finkelstein J. Towards a highly usable, mobile electronic platform for patient recruitment and consent management. Stud Health Technol Inform 2020 Jun 16;270:1066-1070. [doi: 10.3233/SHTI200325] [Medline: 32570545]
27. Golembiewski EH, Mainous 3rd AG, Rahmanian KP, Brumback B, Rooks BJ, Krieger JL, et al. An electronic tool to support patient-centered broad consent: A multi-arm randomized clinical trial in family medicine. Ann Fam Med 2021;19(1):16-23 [FREE Full text] [doi: 10.1370/afm.2610] [Medline: 33431386]

28. Garrison NA, Sathe NA, Antommaria AHM, Holm IA, Sanderson SC, Smith ME, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet Med 2016 Jul;18(7):663-671 [FREE Full text] [doi: 10.1038/gim.2015.138] [Medline: 26583683]
29. Trehearne A. Genetics, lifestyle and environment. UK Biobank is an open access resource following the lives of 500,000 participants to improve the health of future generations. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2016 Mar;59(3):361-367. [doi: 10.1007/s00103-015-2297-0] [Medline: 26753864]
30. Passmore SR, Casper E, Olgin JE, Maguire C, Marcus GM, Pletcher MJ, et al. Setting and motivation in the decision to participate: An approach to the engagement of diverse samples in mobile research. Contemp Clin Trials Commun 2019 Aug 10;16:100428 [FREE Full text] [doi: 10.1016/j.conctc.2019.100428] [Medline: 31463416]
31. Hagiwara N, Berry-Bobovski L, Francis C, Ramsey L, Chapman RA, Albrecht TL. Unexpected findings in the exploration of African American underrepresentation in biospecimen collection and biobanks. J Cancer Educ 2014 Sep;29(3):580-587 [FREE Full text] [doi: 10.1007/s13187-013-0586-6] [Medline: 24243440]
32. Kim P, Milliken EL. Minority participation in biobanks: An essential key to progress. Methods Mol Biol 2019;1897:43-50. [doi: 10.1007/978-1-4939-8935-5_5] [Medline: 30539433]
33. Shavers VL, Lynch CF, Burmeister LF. Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. J Natl Med Assoc 2000 Dec;92(12):563-572. [Medline: 11202759]
34. Kasperbauer TJ, Schwartz PH. Measuring understanding and respecting trust in biobank consent. Am J Bioeth 2019 May;19(5):29-31. [doi: 10.1080/15265161.2019.1587034] [Medline: 31090517]
35. Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A. Broad consent for health care-embedded biobanking: understanding and reasons to donate in a large patient sample. Genet Med 2018 Jan;20(1):76-82. [doi: 10.1038/gim.2017.82] [Medline: 28640237]
36. Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AHM, Aufox SA, et al. Public attitudes toward consent and data sharing in biobank research: A large multi-site experimental survey in the US. Am J Hum Genet 2017 Mar 02;100(3):414-427 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.021] [Medline: 28190457]
37. Antommaria AHM, Brothers KB, Myers JA, Feygin YB, Aufox SA, Brilliant MH, et al. Parents' attitudes toward consent and data sharing in biobanks: A multisite experimental survey. AJOB Empir Bioeth 2018;9(3):128-142 [FREE Full text] [doi: 10.1080/23294515.2018.1505783] [Medline: 30240342]
38. McGregor KA, Ott MA. Banking the future: Adolescent capacity to consent to biobank research. Ethics Hum Res 2019 Jul;41(4):15-22. [doi: 10.1002/eahr.500023] [Medline: 31336038]
39. Kaye J, Moraia LB, Curren L, Bell J, Mitchell C, Soini S, et al. Consent for biobanking: The legal frameworks of countries in the BioSHaRE-EU project. Biopreserv Biobank 2016 Jun;14(3):195-200 [FREE Full text] [doi: 10.1089/bio.2015.0123] [Medline: 27145287]
40. Lipworth W, Kerridge I. Consent to biobank research: Facing up to the challenge of globalization. Am J Bioeth 2015;15(9):58-59. [doi: 10.1080/15265161.2015.1062179] [Medline: 26305758]
41. Warner TD, Weil CJ, Andry C, Degenholtz HB, Parker L, Carithers LJ, et al. Broad consent for research on biospecimens: The views of actual donors at four U.S. medical centers. J Empir Res Hum Res Ethics 2018 Apr;13(2):115-124 [FREE Full text] [doi: 10.1177/1556264617751204] [Medline: 29390947]

Abbreviations

CAB: community advisory board
CTSI: Clinical and Translational Science Institutes
EHR: electronic health record
EngageUC: Engaging University of California Stakeholders for Biorepository Research
IPH: Institute for Precision Health
NIH: National Institutes of Health
OR: odds ratio
UCLA: University of California, Los Angeles
UCON: universal video consent

Edited by G Eysenbach; submitted 09.06.21; peer-reviewed by M Salimi, J Kaye; comments to author 17.08.21; revised version received 23.08.21; accepted 18.09.21; published 08.12.21.

Please cite as:
Lajonchere C, Naeim A, Dry S, Wenger N, Elashoff D, Vangala S, Petruse A, Ariannejad M, Magyar C, Johansen L, Werre G, Kroloff M, Geschwind D
An Integrated, Scalable, Electronic Video Consent Process to Power Precision Health Research: Large, Population-Based, Cohort Implementation and Scalability Study
J Med Internet Res 2021;23(12):e31121
URL: https://www.jmir.org/2021/12/e31121
doi:10.2196/31121
PMID:34889741

Original Paper

Short-Video Apps as a Health Information Source for Chronic Obstructive Pulmonary Disease: Information Quality Assessment of TikTok Videos

Shijie Song[1], PhD; Xiang Xue[2], MSc; Yuxiang Chris Zhao[3], PhD; Jinhao Li[4], BSc; Qinghua Zhu[2], PhD; Mingming Zhao[5], MD

[1]Business School, Hohai University, Nanjing, China

[2]School of Information Management, Nanjing University, Nanjing, China

[3]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

[4]Sino-French Engineer School, Nanjing University of Science and Technology, Nanjing, China

[5]Department of Pulmonary and Critical Care Medicine, Gaochun People's Hospital, Nanjing, China

Corresponding Author:
Mingming Zhao, MD
Department of Pulmonary and Critical Care Medicine
Gaochun People's Hospital
Maoshan Rd 53
Economic Development Area, Gaochun District
Nanjing, 211300
China
Phone: 86 15951973800
Email: mingtze80@163.com

Abstract

Background: Chronic obstructive pulmonary disease (COPD) has become one of the most critical public health problems worldwide. Because many COPD patients are using video-based social media to search for health information, there is an urgent need to assess the information quality of COPD videos on social media. Recently, the short-video app TikTok has demonstrated huge potential in disseminating health information and there are currently many COPD videos available on TikTok; however, the information quality of these videos remains unknown.

Objective: The aim of this study was to investigate the information quality of COPD videos on TikTok.

Methods: In December 2020, we retrieved and screened 300 videos from TikTok and collected a sample of 199 COPD-related videos in Chinese for data extraction. We extracted the basic video information, coded the content, and identified the video sources. Two independent raters assessed the information quality of each video using the DISCERN instrument.

Results: COPD videos on TikTok came mainly from two types of sources: individual users (n=168) and organizational users (n=31). The individual users included health professionals, individual science communicators, and general TikTok users, whereas the organizational users consisted of for-profit organizations, nonprofit organizations, and news agencies. For the 199 videos, the mean scores of the DISCERN items ranged from 3.42 to 4.46, with a total mean score of 3.75. Publication reliability (P=.04) and overall quality (P=.02) showed significant differences across the six types of sources, whereas the quality of treatment choices showed only a marginally significant difference (P=.053) across the different sources.

Conclusions: The overall information quality of COPD videos on TikTok is satisfactory, although the quality varies across different sources and according to specific quality dimensions. Patients should be selective and cautious when watching COPD videos on TikTok.

(J Med Internet Res 2021;23(12):e28318) doi:10.2196/28318

KEYWORDS

COPD; information quality; social media; short-video apps; TikTok

Introduction

Chronic obstructive pulmonary disease (COPD) has become one of the most critical public health problems, which has resulted in huge health care expenditures [1]. In 2015, COPD caused approximately 3.17 million deaths worldwide, which accounted for 5% of global deaths for that year [2]. COPD is notably more severe in low- and middle-income countries than in high-income countries. From 2014 to 2015, approximately 5.9% of adults in the United States were reported to be living with COPD [3], whereas the estimated prevalence was 13.6% in China during the same period [4].

COPD is a preventable and treatable disease, and there are many opportunities to reduce the risk of COPD before and after diagnosis [5,6]. For example, general health consumers could reduce the risk of contracting COPD by quitting smoking and avoiding secondhand smoke [7]. Even after a COPD diagnosis, maintaining a healthy lifestyle can help patients prevent exacerbations and improve well-being [8]. However, limited access to information about COPD has become a significant problem for patients and their caregivers [9]. People living with COPD often report limited knowledge on several points such as the causes of COPD and its consequences [10]. Patients have also often received inadequate guidance about how to recognize the disease, and avoid and manage exacerbations [10,11]. Therefore, effective health communications that inform patients on recommended actions are important for better disease prevention and management, and information communication technologies can play a substantial role in such communication and intervention [12].

Emerging technologies provide great health communication opportunities that can inform and empower COPD patients in regard to disease management [13]. For example, mobile technologies have been extensively used to help COPD patients achieve an early diagnosis [14], make medical appointments [15], promote physical activity [16], consistently self-monitor [17], enhance self-management [18], and reduce COPD exacerbation [12]. Recently, visually rich social media (eg, YouTube, Pinterest) have become popular among COPD patients [19,20]. In general, rich social media have several advantages in health communication. The information can be illustrated, which makes it easier to process and remember than information in the form of plain text [21]. Imagistic health information can elicit affective reactions and motivate consumers' health actions [22]. Some prior studies suggest that COPD patients exposed to visually rich social media are more willing to engage with COPD-related messages [19,23].

Despite the promising potential of any emerging technology, patients' actual use of a technology can be fraught with problems. Some evidence suggests that information quality is one of the most significant concerns when COPD patients seek health information online [24]. Evaluating the quality of online health information sources is not an easy task for most laypeople [25,26], especially for many COPD patients who have low health literacy [27]. According to a survey of COPD patients, approximately half of the respondents reported difficulty distinguishing between high- and low-quality health resources on the internet [27]. Therefore, health care providers should assess the information quality of online COPD information and advise their patients about it.

To the best of our knowledge, the quality of online COPD information in video-based social media has yet to be sufficiently investigated. Stellefson et al [20] reviewed 223 videos employing the instruments of HONcode (Health on the Net) principles; they found that the majority of YouTube COPD videos (69.1%) were of high quality and were mainly created by authoritative sources (eg, health agencies, organizations, news agents, and professionals). Recently, the short-video app TikTok has attracted significant research attention in regard to its health communication. For example, during the COVID-19 pandemic, TikTok's coronavirus-related videos were viewed 93.1 billion times as of July 2020 [28]. The originality, interactivity, and sociable nature of TikTok have given the younger generation a better user experience and sense of engagement while seeking health information [29]. TikTok affords rich information modalities (eg, text, image, audio, and video), and contains ample technology features such as commenting, chatting, following, liking, and live-streaming [30]. These features make the app easier for the general public to use as a source of health information, and penetration and usage of TikTok are also on the rise among some older age groups [31].

We observed that there are many COPD-related videos on TikTok. However, the quality of the information they offer remains unclear. Therefore, to fill this gap, the aim of this study was to evaluate the information quality of COPD videos on TikTok.

Methods

Search Strategy and Data Extraction

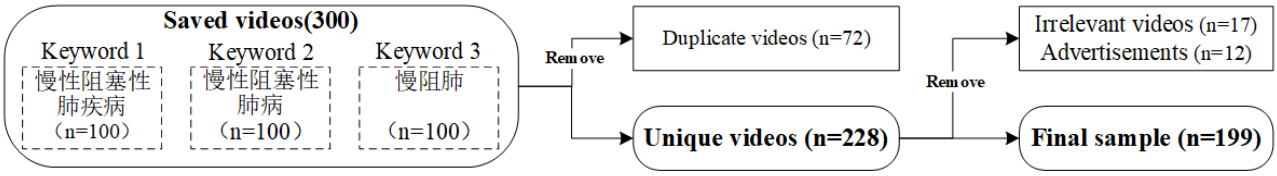
We employed three Chinese words, "慢性阻塞性肺疾病" (chronic obstructive pulmonary disease), "慢性阻塞性肺病" (chronic lung disease related to obstructed airflow), and "慢阻肺" (COPD), to retrieve the relevant COPD videos on TikTok. In its search function, TikTok provides three ways to sort items: "overall ranking," "most recent," and "most likes." Overall ranking is the default mode of sorting recommended by TikTok, which also comprises the other two modes. Given that most users employ the default, we used the overall ranking mode to retrieve the top 100 videos posted from December 6 to December 10, 2020, under each of the three keywords, which resulted in a total of 300 videos. We chose the threshold number of 100 for two reasons. First, TikTok's search function encompasses the consideration of topic relevance; the pertinent COPD videos mostly appear at the top of the result list, and it is hard to observe any relevant videos when the results go beyond 100. Second, most general health consumers apply the "least effort" principle in their online information-seeking activities; thus, they usually view the top search results instead of going very far [32].

To choose the most relevant videos, we removed videos that were (1) duplicates (n=72), (2) not directly related to COPD

topics (n=17), and (3) advertisements (n=12). Finally, a total of 199 videos remained for data analysis (see Figure 1).

We used Microsoft Excel to extract and code the basic information from each video. This included a description of the video; the URL; the upload date; the duration (in seconds); the user ID of the uploader; and the numbers of views, likes, and comments.

Figure 1. Search strategy and video screening procedure.



Instruments

We employed DISCERN as the instrument for assessing the quality of the information in each video. DISCERN is a brief questionnaire designed to help health consumers and researchers assess the quality of health information. The questionnaire contains three parts, devoted to the reliability of a publication (8 items), the quality of information on treatment choices (7 items), and an overall rating of the publication (1 item) [33]. Each of the 16 questions is rated on a scale of 1 (lowest score) to 5 (highest score). We chose DISCERN for several reasons: (1) it is one of the most widely used instruments for studying the quality of health information [34], (2) it has proven to be effective in Chinese contexts [35], and (3) it has proven to be useful for assessing information quality on other video-based platforms (eg, YouTube) [36]. The full instrument is presented in Multimedia Appendix 1.

Moreover, we adopted six questions from Goobie et al [37] to evaluate the video content. These six questions ask to what degree a video addresses the definition of a disease, its signs and symptoms, risk factors, evaluation, management, and outcomes. Each aspect was scored on a three-item scale: not addressed (0 points), partially addressed (1 point), and sufficiently addressed (2 points).

Coding Procedure

Two raters (MZ and YZ) independently evaluated the information quality of each video, employing the DISCERN instrument. Both raters are certified physicians who worked in a local hospital in the division of respiratory disease. The coding procedure contained three stages.

In the first stage, we recorded the basic information of the video publishers (eg, account name, self-description, identity verification status) and the basic information of the videos (eg, publication date, video length, number of likes, number of comments, number of shares). Regarding the video publishers, we categorized the video sources into two main types (ie, individual users and organizational users) by their account names and identity verification status. Further, we identified several subcategories within each source type by their account names, self-description, and video publication records. For example, if a video publisher described themselves as a "scientific writer," we would code the source as an "individual science communicator."

In the second stage, we assessed the video content using the six categories from Goobie et al [37]: the definition of a disease, its signs and symptoms, risk factors, evaluation, management, and outcomes. When we had independently scored the first 30 videos, we found that we were able to reach a consensus on whether a video contains a certain category of content. One rater (MZ) then finished rating the rest of the videos and the other rater (YZ) validated the codes, followed by discussion and resolution of any inconsistencies between the two raters.

In the third stage, we evaluated the information quality by applying the DISCERN instrument. Before starting to score the videos, the two raters first reviewed the official DISCERN scoring instructions, discussing how the tool could be operationalized for evaluating video-based content and making necessary adjustments. Using the DISCERN questions, the two raters scored all videos independently. The overall rating agreement (Cohen κ) was 0.793 (P<.001), which indicated that the rating process had satisfactory interrater reliability. Any between-group analyses regarding DISCERN scores were performed using the Kruskal-Wallis H test in SPSS 22.

Results

Video Characteristics

The COPD videos on TikTok came mainly from two types of sources: individual users and organizational users. Most of the COPD TikTok videos in our sample were contributed by individual users (168/199, 84.4%), whereas a relatively small share was contributed by organizational users (31/199, 15.6%). For each type, we identified three subcategories (see Table 1). Among individual users, health professionals created the most videos, followed by individual science communicators and general TikTok users. Among organizational users, for-profit organizations published the highest number of videos, followed by nonprofit organizations and news agencies.

In the sample, the durations of the videos varied from 10 to 4116 seconds. The videos published by nonprofit organizations were significantly longer than videos from other sources, whereas the videos published by news agencies had the shortest average duration. Videos published by individual science communicators had the second-longest average duration. The average duration for the other sources was under 1 minute (Table 2).

The most recent video was published 22 days prior to data collection, whereas the oldest had been on TikTok for more than 1 year. The 199 videos received a total of 1,696,725 likes and 175,703 comments prior to data collection. The number of

likes varied from 0 to 662,000 for each video, and the number of comments ranged from 0 to 18,000. The videos published by health professionals received the most likes and comments, whereas the videos uploaded by individual science communicators received the least likes and comments. Since their publication, the videos in the overall sample had been shared a total of 167,473 times. The videos uploaded by health professionals were shared the most frequently, whereas the videos created by individual science communicators were shared the least frequently, as shown in Table 2.

Table 1. Characteristics of the sources of chronic obstructive pulmonary disease–related TikTok videos (N=199).

Source	Description	Videos, n (%)
Individual users		
General users	Common TikTok users	30 (15.1)
Individual science communicators	Individuals who participate in general scientific communications, which may include but are not limited to health care domains	39 (19.6)
Health professionals	Individuals who describe themselves as health professionals	99 (49.7)
Organizational users		
For-profit organizations	Private sector organizations	18 (9.0)
Nonprofit organization	Organizations or hospitals operating in the public sector	8 (4.0)
News agencies	Organizations providing news services	5 (2.5)

Table 2. Characteristics of chronic obstructive pulmonary disease–related TikTok videos, by source

Source of videos	Days on TikTok, median (IQR)	Video duration (seconds), median (IQR)	Number of likes, median (IQR)	Number of comments, median (IQR)	Number of shares, median (IQR)
Organizational users					
For-profit organizations	180 (39-297)	59 (37-129)	364 (115-4915)	17 (7-32)	19 (7-133)
Nonprofit organizations	148 (44-225)	158 (74-206)	24 (14-40)	1 (0-2)	1 (0-8)
News agencies	145 (22-147)	34 (18-51)	12 (1-18)	0 (0-0)	3 (1-3)
Individual users					
Health professionals	407 (144-490)	57 (49-59)	1031 (170-3203)	54 (8-161)	58 (12-334)
Individual science communicators	303 (147-491)	85 (45-116)	3 (1-11)	0 (0-1)	0 (0-5)
General users	237 (133-394)	40 (15-59)	19 (9-49)	2 (1-5)	1 (0-7)
Overall	276 (119-478)	57 (43-88)	125 (12-1597)	7 (1-91)	12 (1-118)

Information Quality

The videos covered the six predefined content areas to different degrees, as shown in Figure 2. The results suggested that more than half of the videos (122/199, 61.3%) sufficiently addressed COPD outcomes, whereas only 16 (8.0%) made no mention of outcomes. The second most frequently introduced area concerned the signs or symptoms of COPD, with almost half of the videos (100/199, 50.3%) sufficiently addressing them and 76 (38.2%) only giving partial mentions. Moreover, approximately half of the videos only partially mentioned the topics of COPD evaluation, management, and risk factors. The least frequently mentioned topic was the definition of COPD; only 25 videos (12.6%) sufficiently addressed this and 47 videos (23.6%) did not mention it at all.

We calculated the mean scores for each DISCERN item for the total sample. The scores ranged from 3.42 to 4.46 (mean 3.75). The scores of the eight items measuring publication reliability (items 1-8) varied from 3.42 to 4.46 (mean 3.90). For the seven items assessing the quality of information on treatment choices (items 9-15), the scores ranged from 3.45 to 3.69 (mean 3.56). The remaining item (item 16) measuring overall information quality achieved a mean of 3.85 out of 5 points. We categorized the DISCERN items into three sections according to the original instrument indicated: reliability of the publication, quality of information on treatment choices, and overall rating of the publication (see Table 3).

Videos published by nonprofit organizations had the highest reliability, whereas the videos contributed by general users had the lowest reliability. The differences in publication reliability across the six types of video sources were statistically significant. Regarding the quality of information on treatment choices, the nonprofit organizations provided the highest-quality videos, whereas the general users provided the lowest-quality videos; however, the differences among the various information sources were only marginally significant. With regard to the last item concerning overall rating, the highest-quality videos were created by news agencies and the lowest-quality videos

were generated by for-profit organizations. The overall rating of information quality showed significant differences according to different sources. Overall, the total scores of the entire DISCERN questionnaire across the different sources exhibited significant differences (Table 3).

Figure 2. Percentage of videos addressing each chronic obstructive pulmonary disease topic.

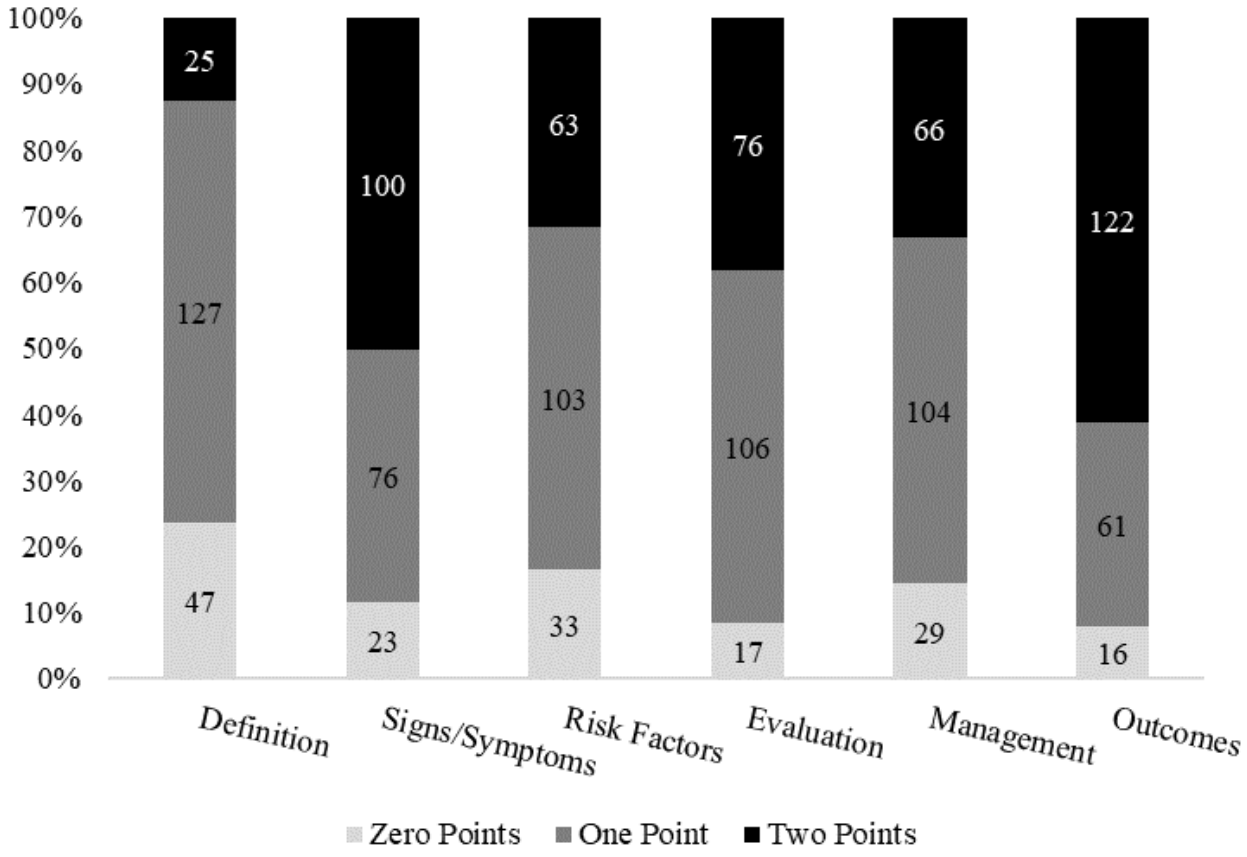


Table 3. DISCERN scores of chronic obstructive pulmonary disease–related TikTok videos by source (N=199).

Category	For-profit organizations (n=18)	Nonprofit organizations (n=8)	News agencies (n=5)	Health professionals (n=99)	Science communicators (n=39)	General users (n=30)	*P* value[a]
Publication reliability							.04
Median	30.5	34.5	34	32	33	30	
Mean (SD)	29.8 (5.9)	34.5 (1.31)	34.1 (1.4)	31.1 (3.7)	31.1 (4.4)	29.7 (5.8)	
Treatment choices							.053
Median	24.3	28	26	25.5	26	23	
Mean (SD)	22.8 (5.7)	28.1 (2.93)	27.3 (3.9)	25.3 (4.2)	25 (4.5)	23.1 (6.1)	
Overall quality							.02
Median	4	4	4	4	4	4	
Mean (SD)	3.4 (0.9)	4.2 (0.65)	4.4 (0.5)	3.8 (0.6)	3.9 (0.7)	3.6 (0.9)	
Total score							.01
Median	60	67	66	61	62	57	
Mean (SD)	56 (11.8)	66.8 (3.8)	65.8 (4.5)	60.4 (7.3)	60 (8.9)	56.5 (11.1)	

[a]*P* values were calculated with the Kruskal-Wallis H test.

Discussion

TikTok as a Health Information Source

Recently, video-based social media platforms and apps have been gaining popularity among patients with chronic conditions [38]. For example, YouTube has become a prominent platform for generating and spreading health-related videos, covering topics related to chronic disease management, including disease prevention, diagnosis, and treatment [39]. Although some recent evidence indicates that TikTok has had vast communication potential during the COVID-19 pandemic [40], the role of TikTok in disseminating chronic disease information remains unclear.

Our results suggest that TikTok could be a promising channel for disseminating COPD information. The 199 videos surveyed in our study have received approximately 1.7 million likes and 176,000 comments since they were published. Given that most of the videos were published within 1 year, the numbers of likes and comments are relatively high. Therefore, we suggest that health care professionals and institutions leverage short-video apps (eg, TikTok) to improve patient education and health communication.

Information Quality

To the best of our knowledge, the quality of online information about COPD is understudied. Specifically in regard to short-video apps, the information quality of COPD videos remains unclear. As one of the earliest studies to tackle this problem, this study yielded results suggesting that the general information quality of COPD videos on TikTok is relatively satisfactory. Although there is a concern that TikTok may differentiate itself from other social media by targeting quirky videos rather than serious professional content [30], our results indicate that the quality of health information found on TikTok, particularly that related to COPD, is acceptable. These results are consistent with a recent survey of coronavirus-related videos on TikTok, which found that the information provided in these videos was generally credible, containing merely 4.3% misinformation [41].

Our findings show that the COPD videos on TikTok more or less touched on all of the preidentified COPD-related content. More than half of the videos sufficiently described the outcomes and the signs and symptoms, and partially addressed the topics of evaluation, risk factors, and management. However, few videos offered definitions of COPD. A possible explanation for this is that TikTok videos are created for a target audience of laypeople. Therefore, introducing aspects of disease management is more important than discussing academic definitions of COPD.

Sources of COPD-Related Videos on TikTok

This study reveals that both individual users and organizational users are engaged in creating COPD videos on TikTok. We identified three specific subcategories for each of these two general categories of sources (ie, individual users and organizational users). Individual users included health professionals, general TikTok users, and individual science communicators. We found that many health professionals were active in communicating COPD knowledge; they contributed almost half of the videos in our sample. Organizational communicators of health information included nonprofit organizations, for-profit organizations, and news agencies. Recently, some researchers have suggested that video-based social media such as TikTok are playing increasingly important roles in providing general health information sought or encountered by consumers [40,41]; therefore, health experts and traditional institutions should harness and leverage social media platforms and apps to better disseminate health information and medical knowledge to the public [41,42]. In general, the findings of this study indicate that health professionals and reputable organizations are actively engaged in communicating health information on TikTok.

In addition, the results suggest that the information quality of COPD videos on TikTok varies according to the source. The videos published by nonprofit organizations had the highest average score for the reliability of publications, while the videos created by general TikTok users had the lowest average score; the differences were statistically significant. The nonprofit organizations and general TikTok users earned the highest and lowest average scores for the quality of information on treatment choices, respectively, although the differences were only marginally significant. For the overall rating of information quality, the news agencies contributed content of significantly higher quality than that of other sources. All of these results are consistent with prior studies in YouTube settings, where organizational users were found to create videos of significantly higher quality than those of individual users [20,37].

Communication performance varied among the sources. In our sample, the videos published by health professionals received the most likes and comments, and were most frequently shared by users. The videos created by individual science communicators received the least likes and comments and were least likely to be shared by users. Interestingly, despite the significant differences in communication performance, we found that the two sources had comparable levels of objective information quality, as assessed by the DISCERN instrument. Prior information credibility studies suggest that credibility depends on the perception of the information recipient, which may not necessarily reflect the objective quality of the information [25,43], and users' credibility perception predicts whether or not they will adopt and share health information on social media [44]. The discrepancy between information quality and communication performance may partially confirm such findings. We suspect that health professionals convey a higher level of expertise and thus generate greater credibility, whereas general science communicators convey low expertise in medicine when they upload videos with wide topic coverage extending beyond the medical domain. We call for future research to investigate the discrepancies between the actual quality of health information and users' credibility perception.

Limitations and Future Research

This study has some limitations. First, we employed only the DISCERN instrument, which was chosen because it has worked well in prior studies that assessed the quality of information in health-related videos. However, there are other instruments

available, such as the JAMA (Journal of the American Medical Association) benchmarks and the HONcode, and future research could expand these investigations by using different instruments. Second, this study only assessed the information quality of Chinese COPD videos. Although the locality did not impact the overall results, the conclusions may not necessarily be generalizable to COPD videos in other languages; therefore, we call for more cross-language comparative studies in the future.

Conclusion

This study investigated the information quality of COPD videos on the world's largest short-video app, TikTok, employing the DISCERN instrument. The results show that both organizational and individual users generate COPD-related content. The overall information quality of the COPD-related videos in the sample was satisfactory, although the quality varied across the different video sources and specific quality dimensions. The results suggest that patients should remain cautious and selective when watching COPD videos on TikTok. Videos from identifiable sources (eg, nonprofit organizations) are much more strongly recommended than those from other nonverified sources. Based on the limitations of this study, we have proposed several directions for future research.

Acknowledgments

We thank Dr Ya Zhang for her contribution to the data analysis and the anonymous reviewers and editors for their helpful comments on the manuscript. The study is supported by the National Science Foundation of China (grant numbers 72174083, 72074112, and 71774083) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant number KYCX19_0022).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The DISCERN questionnaire.
[DOCX File , 13 KB - jmir_v23i12e28318_app1.docx]

References

1. Quaderi S, Hurst J. The unmet global burden of COPD. Glob Health Epidemiol Genom 2018;3:e4 [FREE Full text] [doi: 10.1017/gheg.2018.1] [Medline: 29868229]
2. Chronic obstructive pulmonary disease (COPD). World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd) [accessed 2020-12-29]
3. Sullivan J, Pravosud V, Mannino DM, Siegel K, Choate R, Sullivan T. National and state estimates of COPD morbidity and mortality - United States, 2014-2015. Chronic Obstr Pulm Dis 2018 Oct 12;5(4):324-333. [doi: 10.15326/jcopdf.5.4.2018.0157] [Medline: 30723788]
4. Fang L, Gao P, Bao H, Tang X, Wang B, Feng Y, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. Lancet Respir Med 2018 Jun;6(6):421-430 [FREE Full text] [doi: 10.1016/S2213-2600(18)30103-6] [Medline: 29650407]
5. Li P, Lu X, Teng C, Cai P, Kranis M, Dai Q, et al. The impact of COPD on in-hospital outcomes in patients with Takotsubo cardiomyopathy. Int J Chron Obstruct Pulmon Dis 2020;15:2333-2341. [doi: 10.2147/COPD.S267289] [Medline: 33061351]
6. Li P, Dai Q, Cai P, Teng C, Pan S, Dixon RA, et al. Identifying different phenotypes in takotsubo cardiomyopathy by latent class analysis. ESC Heart Fail 2021 Feb;8(1):555-565. [doi: 10.1002/ehf2.13117] [Medline: 33244882]
7. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. J Med Virol 2020 Oct;92(10):1915-1921 [FREE Full text] [doi: 10.1002/jmv.25889] [Medline: 32293753]
8. de Blok BMJ, de Greef MHG, ten Hacken NHT, Sprenger SR, Postema K, Wempe JB. The effects of a lifestyle physical activity counseling program with feedback of a pedometer during pulmonary rehabilitation in patients with COPD: a pilot study. Patient Educ Couns 2006 Apr;61(1):48-55. [doi: 10.1016/j.pec.2005.02.005] [Medline: 16455222]
9. Nakken N, Janssen DJA, van den Bogaart EHA, Muris JWM, Vercoulen JH, Custers FL, et al. Knowledge gaps in patients with COPD and their proxies. BMC Pulm Med 2017 Oct 30;17(1):136 [FREE Full text] [doi: 10.1186/s12890-017-0481-8] [Medline: 29084521]
10. Hernandez P, Balter M, Bourbeau J, Hodder R. Living with chronic obstructive pulmonary disease: a survey of patients' knowledge and attitudes. Respir Med 2009 Jul;103(7):1004-1012 [FREE Full text] [doi: 10.1016/j.rmed.2009.01.018] [Medline: 19269150]
11. Stellefson M, Paige SR, Alber JM, Chaney BH, Chaney D, Apperson A, et al. Association between health literacy, electronic health literacy, disease-specific knowledge, and health-related quality of life among adults with chronic obstructive pulmonary

disease: cross-sectional study. J Med Internet Res 2019 Jun 06;21(6):e12165 [FREE Full text] [doi: 10.2196/12165] [Medline: 31172962]
12. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble J, Abu Ashour W. The effect of smartphone interventions on patients with chronic obstructive pulmonary disease exacerbations: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 Sep 01;4(3):e105 [FREE Full text] [doi: 10.2196/mhealth.5921] [Medline: 27589898]
13. 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 Jun 04;22(6):e14417 [FREE Full text] [doi: 10.2196/14417] [Medline: 32496198]
14. De Ramón Fernández A, Ruiz Fernández D, Marcos-Jorquera D, Gilart Iglesias V. Support system for early diagnosis of chronic obstructive pulmonary disease based on the service-oriented architecture paradigm and business process management strategy: development and usability survey among patients and health care providers. J Med Internet Res 2020 Mar 17;22(3):e17161 [FREE Full text] [doi: 10.2196/17161] [Medline: 32181744]
15. 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]
16. Bentley CL, Powell L, Potter S, Parker J, Mountain GA, Bartlett YK, et al. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16203 [FREE Full text] [doi: 10.2196/16203] [Medline: 32490838]
17. Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E, et al. Technology-enabled self-management of chronic obstructive pulmonary disease with or without asynchronous remote monitoring: randomized controlled trial. J Med Internet Res 2020 Jul 30;22(7):e18598 [FREE Full text] [doi: 10.2196/18598] [Medline: 32729843]
18. Korpershoek YJG, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JCA. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
19. Paige SR, Stellefson M, Chaney BH, Alber JM. Pinterest as a resource for health information on chronic obstructive pulmonary disease (COPD): a social media content analysis. Am J Health Educ 2015 Jul 06;46(4):241-251. [doi: 10.1080/19325037.2015.1044586]
20. Stellefson M, Chaney B, Ochipa K, Chaney D, Haider Z, Hanik B, et al. YouTube as a source of chronic obstructive pulmonary disease patient education: a social media content analysis. Chron Respir Dis 2014 May;11(2):61-71 [FREE Full text] [doi: 10.1177/1479972314525058] [Medline: 24659212]
21. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related Facebook pages. Ann Behav Med 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
22. Feng B, Malloch Y, Kravitz R, Verba S, Iosif A, Slavik G, et al. Assessing the effectiveness of a narrative-based patient education video for promoting opioid tapering. Patient Educ Couns 2021 Feb;104(2):329-336. [doi: 10.1016/j.pec.2020.08.019] [Medline: 32900605]
23. Paige SR, Stellefson M, Chaney BH, Chaney DJ, Alber JM, Chappell C, et al. Examining the relationship between online social capital and eHealth literacy: implications for Instagram use for chronic disease prevention among college students. Am J Health Educ 2017;48(4):264-277 [FREE Full text] [doi: 10.1080/19325037.2017.1316693] [Medline: 29152031]
24. Paige SR, Stellefson M, Krieger JL, Alber JM. Computer-mediated experiences of patients with chronic obstructive pulmonary disease. Am J Health Educ 2019;50(2):127-134 [FREE Full text] [doi: 10.1080/19325037.2019.1571963] [Medline: 31641399]
25. Song S, Zhang Y, Yu B. Interventions to support consumer evaluation of online health information credibility: A scoping review. Int J Med Inform 2021 Jan;145:104321. [doi: 10.1016/j.ijmedinf.2020.104321] [Medline: 33202372]
26. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
27. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based health information seeking and eHealth literacy among patients living with chronic obstructive pulmonary disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
28. Ostrovsky AM, Chen JR. TikTok and its role in COVID-19 information propagation. J Adolesc Health 2020 Nov;67(5):730 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.039] [Medline: 32873499]
29. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users' intention to continue the use of TikTok. Internet Res 2021 Jun 24;31(6):2120-2142. [doi: 10.1108/intr-10-2020-0593]
30. Wang Y. Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (DouYin). Comput Hum Behav 2020 Sep;110:106373. [doi: 10.1016/j.chb.2020.106373]

31. Nakamura A. TikTok Teens Follow New Stars: Senior Citizens. The Wall Street Journal. 2020. URL: https://www.wsj.com/articles/tiktok-teens-follow-new-stars-senior-citizens-11605724082 [accessed 2021-05-17]
32. Case D, Given L. Looking for information: a survey of research on information seeking, needs, and behavior. UK: Emerald Group Publishing; 2016.
33. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
34. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assn Inf Sci Tec 2015 Apr 29;66(10):2071-2084. [doi: 10.1002/asi.23311]
35. Sun F, Yang F, Zheng S. Evaluation of the liver disease information in Baidu Encyclopedia and Wikipedia: longitudinal study. J Med Internet Res 2021 Jan 18;23(1):e17680 [FREE Full text] [doi: 10.2196/17680] [Medline: 33459597]
36. Aydin M, Aydin M. Quality and reliability of information available on YouTube and Google pertaining gastroesophageal reflux disease. Int J Med Inform 2020 May;137:104107. [doi: 10.1016/j.ijmedinf.2020.104107] [Medline: 32146372]
37. Goobie GC, Guler SA, Johannson KA, Fisher JH, Ryerson CJ. YouTube videos as a source of misinformation on idiopathic pulmonary fibrosis. Annals ATS 2019 May;16(5):572-579. [doi: 10.1513/annalsats.201809-644oc]
38. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021 Sep 01;23(9):e30409 [FREE Full text] [doi: 10.2196/30409] [Medline: 34468327]
39. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics J 2015 Sep;21(3):173-194 [FREE Full text] [doi: 10.1177/1460458213512220] [Medline: 24670899]
40. Comp G, Dyer S, Gottlieb M. Is TikTok the next social media frontier for medicine? AEM Educ Train 2021 Jul;5(3):10.1002/aet2.10532 [FREE Full text] [doi: 10.1002/aet2.10532] [Medline: 34095694]
41. Basch C, Hillyer G, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10:online ahead of print. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
42. Song S, Zhao YC, Yao X, Ba Z, Zhu Q. Serious information in hedonic social applications: affordances, self-determination and health information adoption in TikTok. J Document 2021 Nov 19;ahead-of-print. [doi: 10.1108/JD-08-2021-0158]
43. Zhang Y, Song S. Older adults' evaluation of the credibility of online health information. 2020 Presented at: Conference on Human Information Interaction and Retrieval; March 14-18, 2020; Vancouver, BC p. 358-362. [doi: 10.1145/3343413.3377997]
44. Luo C, Luo X, Schatzberg L, Sia CL. Impact of informational factors on online recommendation credibility: The moderating role of source credibility. Decision Support Syst 2013 Dec;56:92-102. [doi: 10.1016/j.dss.2013.05.005]

Abbreviations

COPD: chronic obstructive pulmonary disease
HON: Health on the Net
JAMA: Journal of the American Medical Association

Edited by G Eysenbach; submitted 05.03.21; peer-reviewed by P Li, P Zou, L Spencer; comments to author 10.05.21; revised version received 23.05.21; accepted 18.11.21; published 20.12.21.

Please cite as:
Song S, Xue X, Zhao YC, Li J, Zhu Q, Zhao M
Short-Video Apps as a Health Information Source for Chronic Obstructive Pulmonary Disease: Information Quality Assessment of TikTok Videos
J Med Internet Res 2021;23(12):e28318
URL: https://www.jmir.org/2021/12/e28318
doi:10.2196/28318
PMID:34931996

Original Paper

An Analysis of US Academic Medical Center Websites: Usability Study

Jonathan James Gale[1], BA; Kameron Collin Black[1], MPH; Joshua David Calvano[1], BS; Edwin Lauritz Fundingsland Jr[1], BA; Deborah Lai[2], BA; Sara Silacci[3], BS; Shuhan He[3], MD

[1]Rocky Vista University College of Osteopathic Medicine, Parker, CO, United States

[2]Division of Psychology and Language Sciences, University College London, London, United Kingdom

[3]Center for Innovation in Digital HealthCare, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Jonathan James Gale, BA
Rocky Vista University College of Osteopathic Medicine
8401 S Chambers Rd
Parker, CO, 80134
United States
Phone: 1 303 373 2008
Email: jonathan.gale@rvu.edu

Abstract

Background: Health care organizations are tasked with providing web-based health resources and information. Usability refers to the ease of user experience on a website. In this study, we conducted a usability analysis of academic medical centers in the United States, which, to the best of our knowledge, has not been previously carried out.

Objective: The primary aims of the study were to the following: (1) adapt a preexisting usability scoring methodology to academic medical centers; (2) apply and test this methodology on a sample set of academic medical center websites; and (3) make recommendations from these results on potential areas of improvements for our sample of academic medical center websites.

Methods: All website usability testing took place from June 1, 2020, to December 15, 2020. We replicated a methodology developed in previous literature and applied it to academic medical centers. Our sample included 73 US academic medical centers. Usability was split into four broad categories: accessibility (the ability of those with low levels of computer literacy to access and navigate the hospital's website); marketing (the ability of websites to be found through search engines and the relevance of descriptions to the links provided); content quality (grammar, frequency of information updates, material relevancy, and readability); and technology (download speed, quality of the programming code, and website infrastructure). Using these tools, we scored each website in each category. The composite of key factors in each category contributed to an overall "general usability" score for each website. An overall score was then calculated by applying a weighted percentage across all factors and was used for the final "overall usability" ranking.

Results: The category with the highest average score was technology, with a 0.82 (SD 0.068, SE 0.008). The lowest-performing category was content quality, with an average of 0.22 (SD 0.069, SE 0.008). As these numbers reflect weighted percentages as an integer, the higher the score, the greater the overall usability in that category.

Conclusions: Our data suggest that technology, on average, was the highest-scored variable among academic medical center websites. Because website functionality is essential to a user's experience, it is justified that academic medical centers invest in optimal website performance. The overall lowest-scored variable was content quality. A potential reason for this may be that academic medical center websites are usually larger in size, making it difficult to monitor the increased quantity of content. An easy way to improve this variable is to conduct more frequent website audits to assess readability, grammar, and relevance. Marketing is another area in which these organizations have potential for improvement. Our recommendation is that organizations utilize search engine optimization techniques to improve their online visibility and discoverability.

(J Med Internet Res 2021;23(12):e27750) doi:10.2196/27750

KEYWORDS

website usability; digital health; health care website; academic medical center; usability testing; web crawler

Introduction

Background

A medical center's website is often the first point of contact with the public; its initial impact is responsible for returning users and attracting new visitors [1,2]. It has the potential to be the first step in improving patient satisfaction as well as attracting new patients [3]. In a time when information is expected to be readily available, medical centers use their websites as a key tool for both patient communication and education [4-6]. The users expect to find current and reliable information on websites that are easily accessible in order to make health-related decisions [7]. With many sources available (eg, WebMD), medical centers are looking to improve their internet presence to better engage with potential consumers [3].

Website Usability

Improving website usability is a noteworthy way in which medical centers can improve their internet presence to attract and retain more users and thus reach a larger audience with accurate and reliable information. Usability goes beyond surface-level design [8] but refers broadly to a product's "user experience," such as ease of navigation or encountered problems within a website [9]. It addresses the question of how easy or pleasing a website is to use (which can influence how many users engage with it), the level of engagement, and a website's ability to achieve other objectives. When users are not able to easily access and use a website, they are unlikely to continue using that given source. Alternatively, improved usability can enhance the reach of a website. For this reason, websites are facing the increasing need to conform to user expectations, desires, and requirements [10,11]. Various industries have established standardized guidelines for accessibility, content, marketing, and technology in order to improve usability [12-14].

Usability for Academic Medical Centers

Studies have sought to apply usability analyses to e-commerce, e-government, mobile news apps, and library websites [15-18]. In health care, other studies have looked at usability for hospitals, children's hospitals, digital health centers, residencies, and cancer center websites [3,19-21]. However, to our knowledge, no studies of usability have been conducted exclusively on academic medical centers in the United States, which included all websites falling under the academic medical centers' domain. Academic medical centers are the intersection of health professional schools, patient care, and academic research. Since an academic medical center comprises numerous institutions that function on their own part to be a part of a greater whole, they play a key role in the advancement of medical care [22]. Web presence is the way in which an academic medical center can demonstrate its advancements in their health professional schools, patient care, and academic research. It is important for these organizations to utilize usability metrics to not only improve user experience, but to represent themselves well as leaders in innovation.

Objectives

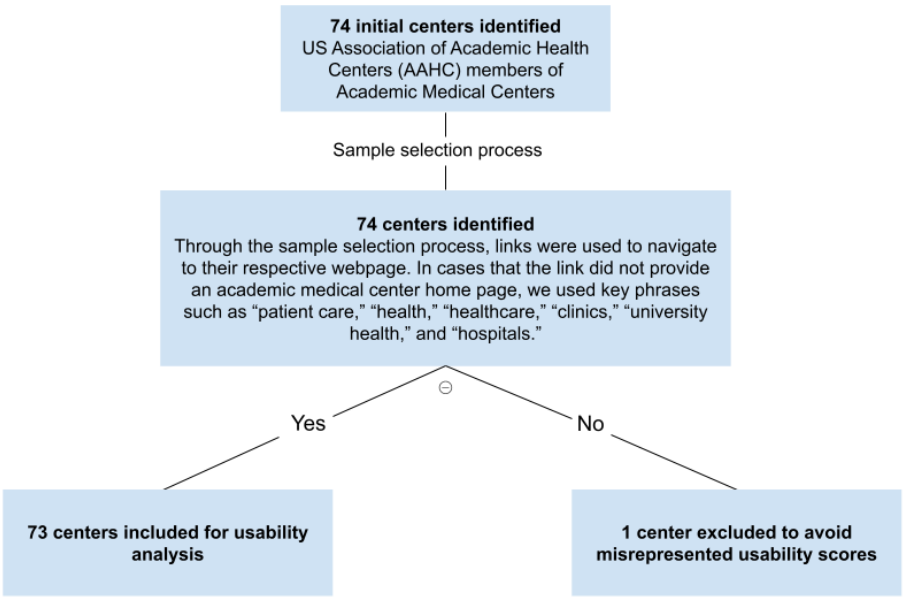
The primary aims of the study were to adapt a usability scoring methodology to academic medical centers; to apply and test this methodology on a sample set of digital health center websites; and to make recommendations from these results on potential areas of improvements for our sample of academic medical center websites.

Methods

Sample Selection

Our focus was on academic health centers in the United States. Indexing every academic medical center in the United States was not in the scope of our review; rather, this study focused on 74 academic medical centers listed on the members page of the Association of Academic Health Centers (AAHC) [23]. A link was provided for each of the academic medical centers listed, and the links were used to navigate to the appropriate academic medical center's page. Some of the links navigated to the affiliated university's website; therefore, terms such as "patient care," "health," "healthcare," "clinics," "university health," and "hospitals," were used to find the appropriate academic medical center's main webpage. One link was removed from analysis, the University of California System, as this link provided a list of academic medical centers in California that were already included in the current analysis. This provided a total of 73 academic medical centers that were used for usability testing (Figure 1).

Figure 1. Sample selection criteria for academic medical center websites.



Overview

All website usability testing took place from June 1, 2020, to December 15, 2020. We began by adapting a methodology developed in previous health care website ranking literature [3,19,24]. We kept the weighted percentages from the previous studies and applied specific formulas to those calculated percentages to create a relative scale for comparing usability scoring (Multimedia Appendix 1). Of note, the previous methodology used a consistent scoring system that used the relative maximum as the reference point of 1 so that every usability test fell between 0 and 1 [24]. However, we found that one of these tests did not follow this format, so our methodology improved the previous testing and applied the same reference range to the values.

Usability was sorted into the following four broad categories in order to ensure quantifiable, objective, and actionable recommendations for the websites: (1) accessibility: the ability of those with low levels of computer literacy to access and navigate the hospital's website; (2) marketing: the ability of websites to be found through search engines and the relevance of descriptions to the links provided; (3) content quality: grammar, frequency of information updates, material relevancy, and readability; and (4) technology: download speed, quality of the programming code, and website infrastructure [3,19].

Each of these categories represents areas of usability in which academic medical centers can communicate more effectively with their audiences. The factors that contribute to each category were originally discussed by Huerta et al [3]. To add to the comprehensive nature of this study, we further defined these categories by utilizing support tools that bolster the credibility and reproducibility of the results of this analysis. These tools were chosen based on their ability to address each of the individual factors in the category criteria and their availability for public ease of use and implementation.

Analysis

All of the websites were analyzed using a set of established usability tools. These tools were chosen based on their ability to specifically assess the individual factors selected. The tools were also chosen based on their ability to meet industry standards for evaluating the selected factor and for their relative ease of use. The process for utilizing each tool was based on the tool's specific instruction manual. The authors collaboratively ensured the proper use of each tool and the proper value to be recorded. There were 2 researchers who collected data, and each underwent specific training on utilizing the suite of analytic tools and data entry. Questions and discrepancies were addressed by the project supervisor as they arose. Factors such as speed, which can vary from second to second, were averaged across 2 separate tools to provide the most accurate values possible. Overall, all the tools were run on a total of 2 computers to minimize as many outlying technology errors as possible. The selected tools can be viewed in Multimedia Appendix 1.

We began building a database of the top-level URLs associated with each website in our data set. The web crawler processes the URL and creates a topographical map of the website, including all its subpages. For instance, a top-level domain, corresponding to a website's home page, may be associated with the URL "www.healthcare.org." A subpage of this center might be a page on the team members and associated with the URL "www.healthcare.org/team." There may be other subpages for specific topics such as the emergency medicine department, the pediatric department, and so forth. Once the web crawler has created a topographical map of a website, that website can then be analyzed for page errors, amount of page content, metadata (ie, titles, keywords, and descriptions), or other preprogrammed factors [25].

Using these usability tools, we scored each website on the four previously mentioned categories. The composite of key factors

in each of those categories contributed to an overall "general usability" score for each website. Lastly, an overall score was calculated by applying a weighted percentage across all factors and used for the final ranking system.

Below, we will describe each of the categories we evaluated and their contributed significance.

Accessibility

Accessibility is a category that refers to how well a website engages a broad audience with varied levels of technical ability, literacy, and disability. This category includes the following components: meta description, functionality, readability, and the overall layout of the website. A meta description refers to the "snippet" page summary presented in a search engine result. Functionality refers to a website's ability to provide content that appeals to a broad range of literacy levels. Functionality also includes features that allow users to access different parts of a website. It is estimated that 43% of American adults have basic or below-basic literacy levels [26]. The usability of assistive technologies such as screen readers and magnifiers for websites is also assessed by the accessibility category [27]. In this study, we utilized tools that apply algorithmic scales to rank website reading difficulty and to determine the grade level required to comprehend a website's content.

Content Quality

The content quality category refers to our assessment of the attributes of the content on a website. This can include the relevancy of the written information to that particular point in time on a specific topic, generated metadata, and the use of the website's multimedia for imagery. For instance, a website dedicated to supplying information on current closed-loop insulin pumps for patients with diabetes may be evaluated on its ability to provide relevant, fact-driven answers to questions that people are seeking answers for (ie, relative costs, ease of use, etc). The multimedia on a website may also be evaluated for issues such as quality (eg resolution) and the available metadata function to add support to the composed content. Content quality also includes written text and may evaluate grammar and spelling.

Marketing

The marketing category refers to our assessment of the discoverability of a website, with a particular emphasis on its search engine results pages (SERP), which refers to the websites presented to users when they search for something online using a search engine such as Google. Higher placement in search results can lead to greater visibility, and SERPs are considered by some to be one of the most important elements of digital marketing. The field of search engine optimization (SEO) deals with optimizing a website to place better in SERPs, and effectively implementing SEO may allow health care websites to uphold a corporate image as industry leaders [28]. However, technical SEO auditing, specifically, was beyond the scope of this study.

Technology

The technology category refers to our assessment of the technical functionality of a website, as opposed to its content. It evaluates the quality of a website's technology and technological design and performance, including its front-end design and user experience as well as back-end coding infrastructure and server management. The front end is what the users of the site view when browsing a website. It also involves the analysis of HTML elements to ensure that the user has an easily navigated layout and that the site can be scalable across devices (ie, computers, mobile phones, and tablets). The back end involves the programming code upon which the website runs. This code and other website components, such as its databases, are stored on servers, which functionally allow people to view websites from their own devices. The servers also affect the speed of the site (eg, how quickly it loads for users), which can play a crucial role in gaining and maintaining users and followers. For instance, a previous study conducted by Google [29] showed that a website that takes longer than 3 seconds to load on a mobile device will lose approximately 53% of its users; problematically, that same study revealed that the average mobile website speed is upwards of 18 seconds.

General Usability

This was a composite of all the metrics from the prior four categories. This category aims to answer the question, "How good is my website?" This metric may serve as a starting point for health care organizations to perform an initial audit of their website to look for areas of improvement.

Overall Usability

An overall usability rank order calculation was included to create a comprehensive evaluation of all major and minor factors across all of the five aforementioned categories. From there, we assigned a percentage weight to create an all-inclusive usability ranking system.

Results

Scores were assigned to all (N=73) academic medical centers found on the AAHC members list [23].

The category with the highest average score was technology, with 0.82 (SD 0.068, SE 0.008). Accessibility was the second highest scoring subcategory, with an average score of 0.77 (SD 0.059, SE 0.007). The third highest scoring subcategory was marketing with an average score of 0.43 (SD 0.066, SE 0.008). The lowest performing category was content quality, with an average of 0.22 (SD 0.070, SE 0.008). The summary statistics across all five categories are presented in Table 1, and a description of the usability tools used in each of the categories can be found in Multimedia Appendix 1.

The overall rankings for the 73 assessed domains for all categories are reported in Multimedia Appendix 2.

Table 1. Academic medical center websites: summary statistics from usability analysis.

Category	Mean (SE)	SD	Minimum	Maximum
Accessibility	0.77 (0.007)	0.059	0.58	0.86
Content quality	0.22 (0.008)	0.069	0.09	0.50
Marketing	0.43 (0.008)	0.066	0.26	0.63
Technology	0.82 (0.008)	0.068	0.66	0.97
General usability	0.62 (0.005)	0.047	0.48	0.71

The top leaders across all usability ranking categories are as follows: (1) accessibility—Duke University; (2) content quality—University of Pittsburgh; (3) marketing—University of Southern California; (4) technology—Rosalind Franklin University of Medicine and Science; and (5) general usability—University of Southern California. The top-performing website in terms of overall usability was that of the University of Southern California.

Discussion

Comparison With Prior Work

This study assessed academic medical center websites utilizing the methodology outlined by Calvano et al [24] in a publication that ranked usability of digital health care center websites. Previous studies enabled Calvano et al to assess website usability trends including that of hospitals, digital health centers, and children's hospitals [3,19,24]. In previous studies, content quality was the highest scoring category. This was postulated to reflect the health care industry's emphasis on providing adequate information for website users. However, investment in content quality is accompanied by a lower investment in other usability categories.

Another major trend in previous studies was that the technology category was the lowest scoring category [3,20,21,24]. Interestingly, when these study methods were applied to academic medical center websites, we found the opposite. Technology was the highest scoring category, and content quality was the lowest. One possible explanation for this is that the significant amount of time elapsed between the studies enabled website technology to be updated. In addition, compared with community hospitals and digital health centers, academic medical centers are larger institutions with more financial capital to invest in website functionality. Academic medical centers also have more expansive websites, which may cause difficulty in monitoring the quality of the large amount of content produced.

With reference to academic medical centers, we assert that an increased level of importance should be placed on monitoring content quality, thereby ensuring a higher standard of information presented to the general public. As medical advancements continue to become more complex and patients become more comfortable with technology use, a larger number of individuals will turn to the internet for assistance in understanding key medical concepts. Therefore, it is of utmost importance that academic medical centers realize their vital role in providing targeted and relevant content for their website users. This includes providing germane, concise answers to common medical questions people may be searching for to obtain new users and maintain the existing ones.

Another difference between previous studies and our findings involved the accessibility category. In the assessment of children's hospital websites, previous studies found accessibility to be the lowest scoring category [19]. However, we found accessibility to be the second highest scoring category. One possible explanation for this might be that academic medical centers better understand the importance of creating content that is built at the proper literacy levels and technical complexity to ensure ease of access for a broad audience. Academic medical centers tend to serve as leaders in the medical community and are often sought for guidance by the general public in areas of medical concern, such as the current COVID-19 pandemic. With this in mind, academic medical centers may understand the importance of creating content that is easily comprehended by a range of website users.

Marketing scores were noted to be lower than originally anticipated [19]. Pertaining to health care, marketing is important to ensure users can easily locate an organization's website within search engine results. It is imperative that academic medical centers employ search engine optimization techniques to enable improved public visibility of their information compared to less authoritative sources.

A specific goal of this research is to promote standardization of website analysis across the health care industry, as it has been neglected previously despite being an important facet of other sectors [12-14]. Recent technological advancements have driven down costs in medicine and increased quality of care [22]. Usability analysis is an important element in this process and enables health care organizations to improve their website presence. With the heightened awareness of technology's importance in health care due to the COVID-19 pandemic, the web presence of academic medical centers is even more essential to ensuring the proper dissemination of health information.

Limitations

This study has several limitations. Not all of the social media accounts were directly accessible from the website, making them more difficult to find through Twitter and Facebook's respective search engines. In most cases, an affiliated Facebook or Twitter page was found. A total of 5 academic medical centers were not associated with a Twitter page, and a total of 2 academic medical centers were not associated with either a Facebook or Twitter page. For the academic medical centers that did not have affiliated Facebook or Twitter pages, this either was because one had not been created for the institution or

because a link was not provided on their webpage and was not able to be found using the social media's search function.

Assessments of a website's speed can vary depending on the time of the day or the day of data collection. This could be due to changes to the website's servers, internet connectivity, or computer hardware. To minimize sampling bias, the same computer on the same network was used when measuring parameters such as website speed.

All of this information was collected over the course of several months; therefore, some measures may have changed since initial evaluation.

The AAHC provides their own definition of an academic medical center, which includes three main components: "...an allopathic or osteopathic medical school; one or more other health profession schools or programs…, and one or more owned or affiliated teaching hospitals or health systems" [30]. In accordance with the definition of an academic medical center provided by the AAHC, a clinic webpage would not be considered an academic medical center. However, for consistency of using the full sample of academic medical centers provided on the AACH website, we decided to use clinic webpages for usability analysis if no other institution could be found that better met the definition of an academic medical center. In total, 4 clinic pages were used in the analysis. For example, in the case of Des Moines University's Osteopathic Medical Center, only the university's clinic page was able to be found via their webpage; therefore, this was the link that was used in the usability analysis.

Our data set was much larger than that of previous studies [21,24] because we analyzed entire institution websites; many of which had between 500,000 and over 1 million URLs. The websites were cut off at 750,000 URLs during web crawling as this was felt to be a sufficient sample size without running into RAM limitations.

Conclusion

As an increasing number of individuals look to the internet for medical information, responsibility will be placed on academic medical centers to maintain high quality websites, given their status as respected sources of current medical information and research. This study offers an analysis of the overall need for improvement in website usability by academic medical centers. The average general usability score was 0.62, showing the necessity for improving usability measures. Academic medical centers may benefit from taking steps to improve various components of their websites in order to reach their audiences. A suggested step is for these organizations to perform periodic usability audits of their websites to identify areas for improvement. Several of these institutions have significant room for improvement of their overall usability, specifically with content quality and marketing, the lowest scoring categories in this analysis. Using content audit tools, institutions can gather data regarding their webpage content and improve upon several factors, including ensuring every webpage has a title of appropriate length, that every webpage has an H1 heading, and that the meta descriptions on the page are concise. Content audits of webpages should be focused on improving the quality of information presented by enhancing aspects such as navigability. Navigability of information can be improved upon by fixing broken web links that are cited on the website, adding alternative text to images, and correctly utilizing keywords on the pages. In terms of marketing, we also recommend confirming that Facebook pages and other social media links are highly visible on home screens. In addition, we recommend ensuring that there is enough content on those social media pages. These and other webpage improvements will lead to the enhanced usability of the webpages for patients and academicians alike, thereby furthering the quality of health care information available online.

Acknowledgments

We would like to thank Dr. Regan Stiegmann and Dr. Cole Zanetti, co-directors of the Digital Health Track at Rocky Vista University, for their support of this project.

Conflicts of Interest

SH is on the advisory board for Covid Act Now and Safeter.App, and is the co-founder and member of the executive board of ConductScience Inc. SH is on the committee for the American College of Emergency Physician Supply Chain Task Force, and received research funding from the Foundation for Opioid Response Efforts (FORE). SH also discloses the following personal fees: Withings Inc, Boston Globe, American College of Emergency Physicians, Maze Eng Inc, ConductScience Inc, Curative Medical Associates, and VIOMed Spa New England. No other disclosures are reported by the authors.

Multimedia Appendix 1
Defined usability factors with their associated percentage weight, assessment tools, impact, and formulas. Used with permission from Calvano et al [24].
[DOCX File , 19 KB - jmir_v23i12e27750_app1.docx]

Multimedia Appendix 2
Academic medical center websites and category scores.
[DOCX File , 32 KB - jmir_v23i12e27750_app2.docx]

References

1. Alpay LL, Overberg RI, Zwetsloot-Schonk B. Empowering citizens in assessing health related websites: a driving factor for healthcare governance. IJHTM 2007;8(1/2):141. [doi: 10.1504/ijhtm.2007.012103]
2. Revere L, Robinson L. How healthcare organizations use the Internet to market quality achievements. J Healthc Manag 2010;55(1):39-49. [Medline: 20210072]
3. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
4. Coile RC. E-health: reinventing healthcare in the information age. J Healthc Manag 2000;45(3):206-210. [Medline: 11066968]
5. Randeree E, Rao HR. E-health and assurance: curing hospital websites. Int J Electron Healthc 2004;1(1):33-46. [doi: 10.1504/IJEH.2004.004653] [Medline: 18048202]
6. Wen K, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010 Dec 18;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
7. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
8. Griffiths KM, Christensen H. Website quality indicators for consumers. J Med Internet Res 2005 Nov 15;7(5):e55 [FREE Full text] [doi: 10.2196/jmir.7.5.e55] [Medline: 16403719]
9. Otis M, Zhu J, Mustafa-Kutana S, Bernier AV, Ma Shum J, Soros Dupre AA, et al. Testing Usability and Feasibility of a Mobile Educator Tool for Pediatric Diabetes Self-Management: Mixed Methods Pilot Study. JMIR Form Res 2020 May 01;4(5):e16262 [FREE Full text] [doi: 10.2196/16262] [Medline: 32356773]
10. Mittler JN, Volmar KM, Shaw BW, Christianson JB, Scanlon DP. Using websites to engage consumers in managing their health and healthcare. Am J Manag Care 2012 Sep;18(6 Suppl):s177-s184. [Medline: 23286713]
11. Liang C, Chen H. A study of the impacts of website quality on customer relationship performance. Total Quality Management & Business Excellence 2009 Sep;20(9):971-988. [doi: 10.1080/14783360903181784]
12. Oermann MH, Lowery NF, Thornley J. Evaluation of Web sites on management of pain in children. Pain Manag Nurs 2003 Sep;4(3):99-105. [doi: 10.1016/s1524-9042(03)00029-8] [Medline: 14566707]
13. Oermann MH, Lesley ML, VanderWal JS. Using Web sites on quality health care for teaching consumers in public libraries. Qual Manag Health Care 2005;14(3):188-195. [doi: 10.1097/00019514-200507000-00008] [Medline: 16027597]
14. Oermann MH, McInerney SM. An evaluation of sepsis Web sites for patient and family education. Plast Surg Nurs 2007;27(4):192-196. [doi: 10.1097/01.PSN.0000306184.95812.8a] [Medline: 18165726]
15. Kuan H, Bock G, Vathanophas V. Comparing the effects of usability on customer conversion and retention at e-commerce websites. 2005 Jan Presented at: Proc Annu Hawaii Int Conf Syst Sci; 2005; Waikoloa Village p. 174. [doi: 10.1109/hicss.2005.155]
16. Huang Z, Benyoucef M. Usability and credibility of e-government websites. Government Information Quarterly 2014 Oct;31(4):584-595. [doi: 10.1016/j.giq.2014.07.002]
17. Jeong W, Jung Han H. Usability study on newspaper mobile websites. OCLC Systems & Services 2012 Oct 26;28(4):180-198. [doi: 10.1108/10650751211279120]
18. Vasantha Raju N, Harinarayana NS. An analysis of usability features of library web sites. Ann Libr Inf Stud. 2008. URL: http://nopr.niscair.res.in/bitstream/123456789/7680/4/ALIS%2055%282%29%20111-122.pdf [accessed 2021-11-26]
19. Huerta TR, Walker DM, Ford EW. An Evaluation and Ranking of Children's Hospital Websites in the United States. J Med Internet Res 2016 Aug 22;18(8):e228 [FREE Full text] [doi: 10.2196/jmir.5799] [Medline: 27549074]
20. Huerta TR, Walker DM, Ford EW. Cancer Center Website Rankings in the USA: Expanding Benchmarks and Standards for Effective Public Outreach and Education. J Cancer Educ 2017 Jun;32(2):364-373. [doi: 10.1007/s13187-015-0931-z] [Medline: 26472325]
21. Fike J, Fundingsland E, Calvano J, Raja A, He S. 204 Website Usability Analysis of United States Emergency Medicine Residencies. Annals of Emergency Medicine 2020 Oct;76(4):S79. [doi: 10.1016/j.annemergmed.2020.09.217]
22. DePasse JW, Chen CE, Sawyer A, Jethwani K, Sim I. Academic Medical Centers as digital health catalysts. Healthc (Amst) 2014 Sep;2(3):173-176. [doi: 10.1016/j.hjdsi.2014.05.006] [Medline: 26250503]
23. Members. Association of Academic Health Centers. URL: https://www.aahcdc.org/About/Members [accessed 2020-12-21]
24. Calvano JD, Fundingsland EL, Lai D, Silacci S, Raja AS, He S. Applying Website Rankings to Digital Health Centers in the United States to Assess Public Engagement: Website Usability Study. JMIR Hum Factors 2021 Mar 29;8(1):e20721 [FREE Full text] [doi: 10.2196/20721] [Medline: 33779564]
25. Devi RS, Manjula D, Siddharth RK. An Efficient Approach for Web Indexing of Big Data through Hyperlinks in Web Crawling. ScientificWorldJournal 2015;2015:739286 [FREE Full text] [doi: 10.1155/2015/739286] [Medline: 26137592]
26. Kutner M, Greenberg E, Jin Y, Boyle B, Hsu Y, Dunleavy E. Literacy in Everyday Life: Results from the 2003 National Assessment of Adult Literacy. Natl Cent Educ Stat 2007:490.

27. Ismail A, Kuppusamy KS, Nengroo AS. Multi-tool accessibility assessment of government department websites: a case-study with JKGAD. Disabil Rehabil Assist Technol 2018 Aug;13(6):504-516. [doi: 10.1080/17483107.2017.1344883] [Medline: 28766367]
28. Batraga A, Rutitis D. Corporate identity within the health care industry. ecoman 2012 Dec 07;17(4):1545-1551. [doi: 10.5755/j01.em.17.4.3027]
29. Find out how you stack up to new industry benchmarks for mobile page speed. Think with Google. URL: https://www.thinkwithgoogle.com/marketing-resources/data-measurement/mobile-page-speed-new-industry-benchmarks/ [accessed 2020-12-21]
30. Academic Health Centers. Association of Academic Health Centers. URL: https://www.aahcdc.org/About/Academic-Health-Centers [accessed 2020-12-20]

Abbreviations

AAHC: Association of Academic Health Centers
SEO: search engine optimization
SERP: search engine results page

Edited by R Kukafka; submitted 04.02.21; peer-reviewed by M Nitsch, D Walker; comments to author 26.02.21; revised version received 19.04.21; accepted 08.11.21; published 21.12.21.

Please cite as:
Gale JJ, Black KC, Calvano JD, Fundingsland Jr EL, Lai D, Silacci S, He S
An Analysis of US Academic Medical Center Websites: Usability Study
J Med Internet Res 2021;23(12):e27750
URL: https://www.jmir.org/2021/12/e27750
doi:10.2196/27750
PMID:34932015

Original Paper

How to Implement Digital Services in a Way That They Integrate Into Routine Work: Qualitative Interview Study Among Health and Social Care Professionals

Janna Nadav[1,2], RN, MSc; Anu-Marja Kaihlanen[1], RN, PhD; Sari Kujala[3], PhD; Elina Laukka[4], RN, MSc; Pirjo Hilama[5], RN, MSc; Juha Koivisto[1], PhD; Ilmo Keskimäki[1,2], MD, PhD; Tarja Heponiemi[1], PhD

[1]Finnish Institution for Health and Welfare, Helsinki, Finland

[2]Department of Health and Social Care Systems, Tampere, Finland

[3]Aalto University, Espoo, Finland

[4]Oulu University, Oulu, Finland

[5]South Savo Social and Health Care Authority, Mikkeli, Finland

Corresponding Author:
Janna Nadav, RN, MSc
Finnish Institution for Health and Welfare
PO Box 30
Mannerheimintie 166
Helsinki, FI-00271
Finland
Phone: 358 2952480
Email: janna.nadav@thl.fi

Abstract

Background: Although the COVID-19 pandemic has significantly boosted the implementation of digital services worldwide, it has become increasingly important to understand how these solutions are integrated into professionals' routine work. Professionals who are using the services are key influencers in the success of implementations. To ensure successful implementations, it is important to understand the multiprofessional perspective, especially because implementations are likely to increase even more.

Objective: The aim of this study is to examine health and social care professionals' experiences of digital service implementations and to identify factors that support successful implementations and should be considered in the future to ensure that the services are integrated into professionals' routine work.

Methods: A qualitative approach was used, in which 8 focus group interviews were conducted with 30 health and social care professionals from 4 different health centers in Finland. Data were analyzed using qualitative content analysis. The resulting categories were organized under the components of normalization process theory.

Results: Our results suggested 14 practices that should be considered when implementing new digital services into routine work. To get professionals to understand and make sense of the new service, (1) the communication related to the implementation should be comprehensive and continuous and (2) the implementation process should be consistent. (3) A justification for the service being implemented should also be given. The best way to engage the professionals with the service is (4) to give them opportunities to influence and (5) to make sure that they have a positive attitude toward the service. To enact the new service into professionals' routine work, it is important that (6) the organization take a supportive approach by providing support from several easy and efficient sources. The professionals should also have (7) enough time to become familiar with the service, and they should have (8) enough know-how about the service. The training should be (9) targeted individually according to skills and work tasks, and (10) it should be diverse. The impact of the implementation on the professionals' work should be evaluated. The service (11) should be easy to use, and (12) usage monitoring should happen. An opportunity (13) to give feedback on the service should also be offered. Moreover, (14) the service should support professionals' work tasks.

Conclusions: We introduce 14 practices for organizations and service providers on how to ensure sustainable implementation of new digital services and the smooth integration into routine work. It is important to pay more attention to comprehensive and continuing communication. Organizations should conduct a competence assessment before training in order to ensure proper alignment. Follow-ups to the implementation process should be performed to guarantee sustainability of the service. Our findings

from a forerunner country of digitalization can be useful for countries that are beginning their service digitalization or further developing their digital services.

(J Med Internet Res 2021;23(12):e31668) doi:10.2196/31668

KEYWORDS

digital services; implementation; health and social care professionals; integration; normalization process theory; interview; social work; health care; focus groups

Introduction

Background

The number of new digital services has been rapidly growing in the health care setting in recent years. Moreover, the COVID-19 pandemic has significantly boosted the implementation of digital services with unprecedented speed and influence [1,2]. The pandemic has also taken the usefulness and potentials of digital services to a whole new level. In doing so, it has also provided an opportunity to add these services to health care systems in the long term [3]. However, despite the pandemic having recently favored the transition to digital solutions [4], digitalization in health care has been slow and complicated, even though major investments have been made [5].

Implementations of digital services tend to fail more often in the health care setting than in other settings because the environment is complex, and therefore the integration into practice is difficult and slow [6-8]. Failure to implement may even lead to a reduction in quality, safety, and efficiency in care [9]. According to the World Health Organization (WHO), guidance for digital health, research, and assessment of the impact of digital service implementations on health care are essential [4]. It is important to identify barriers and success factors when implementing new digital services [10].

A recent systematic review provided a list of barriers and success factors for the implementation of digital services from the organizational point of view [11]. The most mentioned barrier was limited knowledge of the service, and the most mentioned success factor was the services' ease of use. Resistance from professionals is a major problem for organizations, and therefore it is important to understand their point of view [12]. Health and social care professionals are key influencers in the implementation [13-15] because their attitudes and behaviors influence patients' capacity to use services and their trust in the services [15,16].

Previous findings show that implementation should include users' participation at different implementation phases, using champions or other key staff, providing sufficient training and support, and monitoring the use of the system at the early stages of implementation [13,15,17-19]. However, these studies have mainly focused on examining implementation in certain groups of professionals [11,20] or a single digital service in a particular environment [21-25]. There are relatively few recent studies about health and social care professionals' experiences of the implementation of the digital services from a multiprofessional perspective, especially now when the number of implementations has grown and different professional groups have more practical experience about the implementations.

Moreover, now when the COVID-19 pandemic has further accelerated the adoption of the implementations, it is increasingly important to understand how these solutions are integrated into routine work [4]. Digital services are used for varying tasks and purposes, and creating a current overview of the experiences of professionals over a wide range of digital service implementations rather than focusing on 1 specific implementation would be of benefit. In addition, because Finland is the leading country for the third year in a row in digitalization according to the International Digital Economy and Society Index (I-DESI) [26], perceptions from Finnish professionals about the implementations can provide valuable information for many organizations that are further developing their digital services and systems.

Objectives

The aim of this study is to examine health and social care professionals' experiences of digital service implementations and to identify factors that support successful implementations and should be considered in the future to secure that the services are integrated into professionals' routine work. Normalization process theory (NPT) was used as an analytic framework [27]. This enables an understanding of how digital services can be normalized into professionals' routine work and workflow.

Methods

Design and Settings

A qualitative descriptive design with group interviews was used. The design was chosen because it allows information to be collected directly from those who are experiencing the phenomenon under investigation, as in this study from health and social care professionals [28]. We sought to examine health and social care professionals' experiences with successful implementations by asking them to share hindering and facilitating factors and what should be considered to achieve a successful implementation in routine work. This study followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines [29].

The interviews were conducted at 4 different health centers located in different parts of Finland. These health centers were selected because they were forerunners in digitalization as they had adopted new ways of working and operating digitally before and during the COVID-19 pandemic. Each of these health centers were either pilots or pioneers in the implementation of various digital services (eg, digital symptom questionnaires, self-management instructions, and remote health care

appointments). More examples of the level of each health center's digitalization and the services it provides can be found in Multimedia Appendix 1.

In Finland, health care services are divided into primary health care and specialized medical care. Municipalities (local governments) operate health centers, which represent citizens' first points of contact in public health care. Health centers provide a wide scope of primary health care services, such as general practice outpatient care, maternity and child health clinics, health promotion, oral health care, medical rehabilitation, home nursing, and laboratory and basic imaging services, as well as community hospital care. In some health centers, services may also include some specialized care, such as mental health and substance abuse services [30].

Participants

The participants were health and social care professionals (N=30) working in 4 health centers (Table 1). They were purposely selected by asking clinic managers to recruit volunteers. The inclusion criteria were that the participants had to be health and social care professionals who did client work at health centers that have implemented digital services into their work, and therefore they had recent experience with digital service implementations. Those professionals who had indicated their willingness to participate were contacted via email with information about the study and were asked their willingness to participate in the interviews. All participants provided written informed consent.

Table 1. Demographic characteristics of interviewed professionals (N=30).

Category and variables	Value, n (%)
User group	
Registered nurses	8 (27)
Public health nurses	5 (17)
Practical nurses	7 (23)
Physicians	5 (17)
Social workers	3 (10)
Social counselor	1 (3)
Digital counselor	1 (3)
Gender	
Male	3 (10)
Female	27 (90)
Age (years)	
<30	5 (17)
30-40	11 (36)
41-50	10 (33)
51-60	4 (14)
Career in this organization (years)[a]	
<1	1 (3)
1-5	15 (50)
5-10	3 (10)
10-20	5 (17)
≥20	3 (10)
Career in total (years)[a]	
<1	1 (3)
1-5	7 (23)
5-10	6 (20)
10-20	6 (20)
≥20	5 (17)

[a]Not all of the participants answered this question.

Data Collection

The data were collected with 8 semistructured focus group interviews. The focus group interview method is often used as a qualitative approach when the aim is to obtain data from a purposely selected group of individuals [31], for example, multiprofessional groups such as those in our study. We conducted 2 interviews in each organization. In each focus group, there were 4-6 participants from different professional groups, including physicians, registered nurses, public health nurses, practical nurses, social workers, social counselors, and digital counselors (Table 1). Five of the interviews were conducted face to face, but because the COVID-19 pandemic got worse in Finland, the rest of the interviews were conducted remotely using the Microsoft Teams application. The interviews were performed by 3 interviewers (authors JN, A-MK, and EL) from a research team with previous experience in conducting qualitative interview studies and experiences with digital service implementations.

The questions in the interview guide were based on the literature [11,15,17,18] and defined in collaboration with the research team (Multimedia Appendix 2). The interview guide included questions about the professionals' experiences with digital service implementations, such as which factors facilitated or hindered the implementations and what kind of suggestions the professionals had for future implementations to ensure that services integrate into routine work. Demographic questions related to age, gender, education, and working years in the current organization and total years of working.

The questions in the interview guide were tested in a pilot interview with 1 health care professional by using the Microsoft Teams application. In addition, participants in the first focus group were asked to rate the understandability and relevance of the questions. No changes were required to the interview guide, so the pilot interview was included in the study with the consent of the interviewees. With the permission of the participants, the interviews were recorded and then transcribed by a transcription company. The transcribed text was generated on 168 pages with a line spacing of 1.15, 11-point font, and the font style Verdana. The duration of the interviews ranged from 41 to 79 min, and the total duration of all the interviews was 501 min.

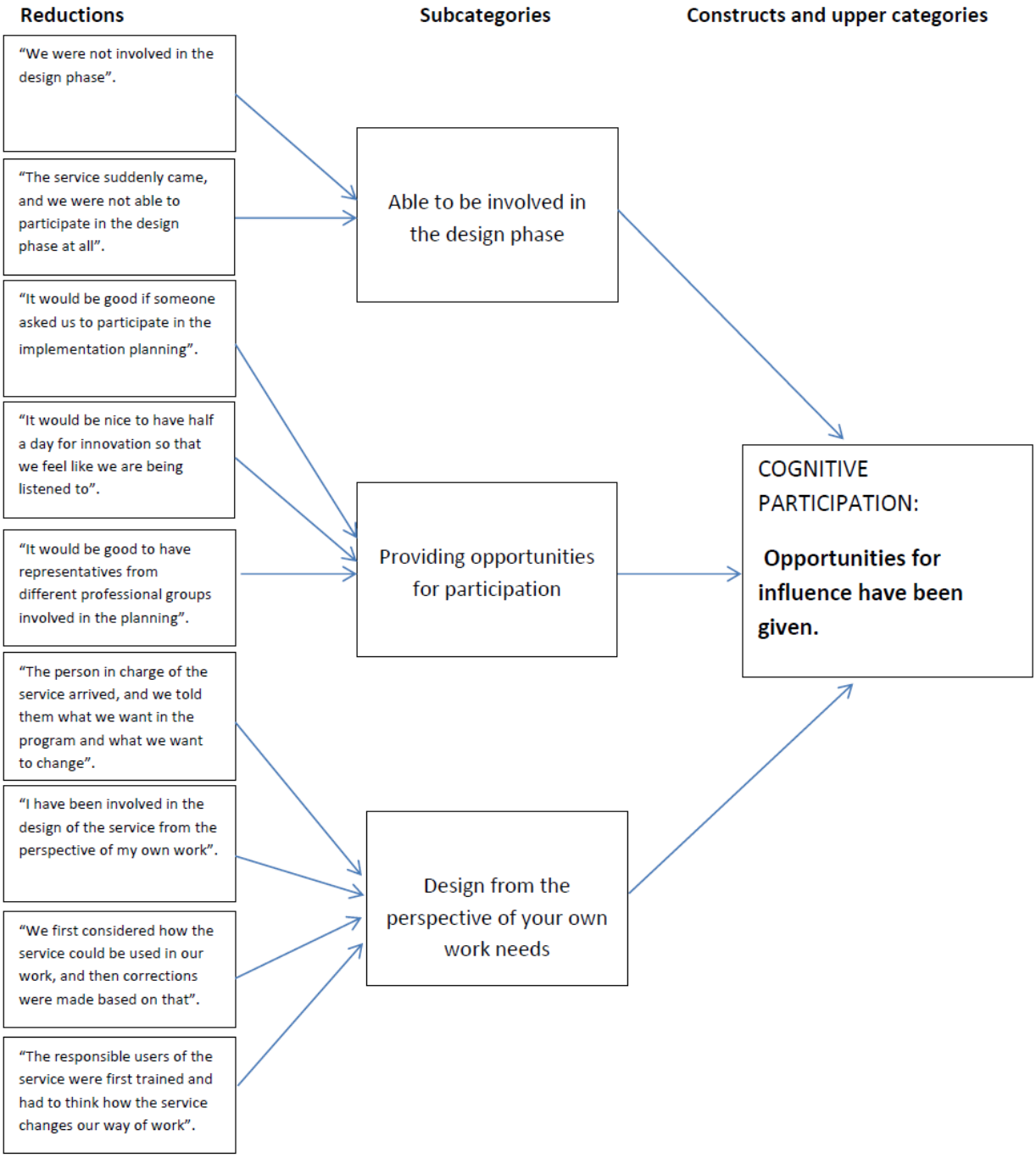
Data Analysis

We discussed the data saturation after the sixth focus group [32], recognizing that responses began to replicate one another and professionals had similar types of experiences with the implementation. The data were analyzed by using content analysis with an inductive-deductive approach [33]. First, 1 researcher (JN) read all the transcribed interviews a couple of times to form a preliminary image of the data. Then, all the expressions that responded to the aim of the study were extracted from the text and formed into codes (n=224) using ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH). The codes were reductions of professionals' thoughts. Subcategories were then formed by grouping codes with similar content, which were then formed into upper categories. At this point, 2 other researchers (authors A-MK and SK) looked at the coding and the formed subcategories and upper categories, and discussions were held to reach an agreement on the content and names. The deductive method was then followed, in which the upper categories were divided into 4 different components according to the NPT framework (Figure 1).

We chose the NPT framework because it is a commonly used framework in implementation studies describing how new technologies and other complex interventions are normalized into routine work in health care settings [34]. It is an action theory, which means that it is concerned with explaining what people do rather than their attitudes or beliefs. Thus, it allowed us to understand the key actions that either promote or inhibit the implementation and integration of the services into professionals' routine work [35]. Moreover, we chose this theory for our study because it can be used to describe and judge the potential of the implementation, but it also has the ability to design and improve complex interventions [27]. We used all 4 of the NPT's components to obtain an overview of the implementations, contrary to previous studies, which mainly focused on just some of the components [36].

The first component (coherence) seeks to explain how the new service changes work and what are the aims and benefits of the service [27]. The second component (cognitive participation) focuses upon the work undertaken to engage people using the service and get them to buy into it [36]. The third component (collective action) refers to work that enables the implementation to happen [36]. It requires the organization to be supportive and people to have the necessary skills and training to perform the tasks associated with the digital services [27]. The fourth component (reflexive monitoring) includes questions such as whether people try to change the practice to fit their work, how they value the new digital services, and what effects the service has on peoples' work [27].

Figure 1. An example of the development of the content analysis process.



Ethics Statements

The Research Ethics Committee approval (THL/2304/6.02.01/2020) was applied for ethical support from the Finnish Institute for Health and Welfare. The data collected in the study were treated confidentially, and the results are reported in a way that does not identify an individual respondent.

Results

Major Findings

We identified 14 practices that, based on the experiences of the professionals, support successful implementation. A detailed description of how the different practices are distributed under the components of the NPT is given in Table 2. The quotation abbreviation meanings are as follows: I=interview and P=participant.

Table 2. Good practices when implementing digital services based on NPT[a] components (n=224).

NPT component and good practices[b]	Mentions[c], n (%)
Coherence (sense-making work): how professionals understand and make sense of the new service (n=38, 17%)	
Communication is comprehensive and continuous: • The information is multichannel • The service presentation reaches everyone. • The service has been informed	28 (13)
The implementation process is consistent: • The implementation process needs to be clear • There is enough time to get ready for the implementation	7 (3)
The use of the service is justified: • The reasons for using the service are given	3 (1)
Cognitive participation (relational work): how professionals engage and participate in the service (n=20, 9%)	
Opportunities for influence have been given: • Professionals are able to be involved in the design phase • Professionals are provided opportunities for participation • The design is from the perspective of the professionals' own work needs	15 (7)
The attitude toward the service is positive: • Previous positive experiences toward eHealth implementations occur • Professionals show interest in the service • Professionals accept the need for the implementation	5 (2)
Collective action (enacting work): the work that individuals (professionals) and organizations have to do to enact the new service (n=136, 60%)	
Support is provided from several fast and efficient sources: • Support is given • The support model is clear • Support is close and easily accessible • Support is given by champions • Support is received from the work community itself • Faster/more efficient sources of support are needed	39 (17)
Sufficient time is provided for familiarization with the service: • Time is provided for familiarization with the service • Independent information retrieval and usage learning are required • The service must be learned alongside the work • A demo version is needed to practice before deployment	36 (16)
Enough knowledge of the service is provided: • Coworkers teach each other with sufficient skills • There are no shortcomings in basic technical skills • There is a need for nonstop training • Sufficient and clear information about the use of the service to support its use is provided • Training is systematically planned	31 (14)
The training is targeted according to work tasks and competence: • There is a need for a competence survey • Training is targeted according to professionals' work tasks • Training is targeted according to professionals' skill level/needs	18 (8)
Various teaching methods are provided: • Versatile teaching methods are available • Good and clear written instructions are provided • Video training is needed to support learning	12 (5)
Reflexive monitoring (appraisal work): how professionals reflect on or appraise the effects of the services (n=30, 13%)	

NPT component and good practices[b]	Mentions[c], n (%)
The service is easy to use: • The assumed heavy usability of the program prevents successful deployment • Experiences with poor usability affect introducing new programs • The service is easy to use • The service has no functional weaknesses	12 (5)
Usage monitoring takes place: • There is continuity of deployment monitoring and evaluation	9 (4)
Giving feedback on the service is possible: • The feedback channel is known • Sending the feedback forward is smooth	5 (2)
The service supports work tasks: • The service is perceived as useful	4 (2)

[a]NPT: normalization process theory.

[b]The categories inside the NPT's components are sorted in the order in which the participants mentioned the most.

[c]How many times the participants mentioned this category.

Coherence: How Do Professionals Understand and Make Sense of the New Service?

The participants considered that the implementation should be prepared by *comprehensive and continuous communication*. They experienced that the communication had failed while announcing that information about the upcoming implementation happened unexpectedly and therefore the presentations had not reached everyone. For future implementations, participants suggested that the communication be done through different communication channels so that it reaches as many employees as possible and avoids uncertainties in the implementation of the service.

> *Implementation is considered successful when the communication is multichannel, which includes a video, a brochure, and a physical person to talk about the service.* [I7, P3]

According to the participants, for digital service implementation to be successful, *the implementation process must be consistent*. The implementation process needs to be clear, and the professionals should have enough time to get ready for the upcoming implementation. The participants had previous experiences with how implementations had taken place on a tight schedule, and thereby a lot of ambiguity had been associated with the process. Some of them had previous experiences where they were first told that the service was going to be implemented but later it was canceled and they attended unnecessary trainings. They started to lose their trust in future implementations and changes.

> *You will lose trust for any change when you experience unclear implementation experiences.* [I2, P1]

In addition, the participants also felt that it is important *to provide a good justification for using the service*. They felt that telling them why the service was implemented and why it needs to be used gave them the motivation to use it in their routine work. In addition, accepting the digitalization and understanding it as a mandatory way of working were contributing factors. Some of the participants even felt that the time is apt for digitalization and that they must simply go with the flow. They also considered it important that the benefits of using the digital service (from the perspective of their own work) were emphasized.

> *Implementation is enabled by the fact that the employee [themselves] perceive the service as useful and good in [their] own work.* [I3, P1]

Cognitive Participation: How Do Professionals Engage With and Participate in the Service?

The participants noted that in good implementation, champions who have an interest and additional training in the service should be involved in implementation from the beginning. They also highlighted that it is important to give everyone, not only the champions, *an opportunity to influence and participate* by giving insight for the service of one's own work needs. The participants suggested that implementation be facilitated by involving professionals from different professional groups in the design phase so that everyone would have a voice. In addition, the participants mentioned that it is important that they have been given the option to use working hours when participating in the design phase of digital services.

> *Usually, we have no time to innovate. It would be nice if there was, for example, half a day for development, which would give the staff a voice.* [I2, P3]

Ensuring that the professionals have a *positive attitude toward the service* enables successful implementation according to the participants. If the professionals had prejudices or negative attitudes toward the digital services, they were reluctant to use them. The participants expressed that usually these negative emotions were consequences of a lack of involvement in the design phase. Prejudices were also often related to the usability of the service. It was feared to be too difficult to use or the

professionals simply had bad experiences with previous implementations.

> *Prejudices can negatively affect successful implementation because you may have some perception or fear that the service [that] is being implemented is too difficult to use.* [I7, P3]

Collective Action: The Work That Professionals and Organizations Must Do to Enact the New Service

The participants pointed out that *support should be available from a variety of sources*. Lack of support or the support model being unclear was found to hinder successful implementation. They noted that everyone should know how the support is provided, whom to contact in the case of a problem, and where to find all the related contact information. The participants had 2 views on the source of support. Some believed that support required a person physically present who could be approached quickly in problem situations. Especially after the first weeks of implementation, this was considered important. For some, remote support was thought to be adequate, as this would allow support to remotely connect the user's computer, if necessary, thus quickly supporting the user in the event of a problem. Nevertheless, what was considered to be the most important thing was that the support should be close and readily available.

> *The person providing the support does not have to be physically present, but it would be good to have support remotely that is easily and quickly available so that [they] can remotely show how to do these things with the service.* [I8, P3]

In addition to the support provided by the organization, support was often received from one's own workplace, most often from one's own colleagues. Some of the participants had experiences of champions supporting the use of the service. In addition to receiving support from their own colleagues, some of the participants also received support from their own supervisors. However, only a few considered it to be important that the supervisors support employees in the implementation process. For them, the attitude of supervisors toward digital services was the more substantial contributing practice to successful implementation.

Most of the participants mentioned that there should be *enough time to get familiar with the service.* They experienced that they had not been provided enough time to get to know the service but had to learn to use it alongside their work while the patient was at the reception. They suggested that by having a demo version, they would get an opportunity to practice the service independently or with colleagues before using it with patients. It would also give them the chance to practice it in peace and when it suits them best. However, some of the participants saw that the use of the service would be learned over time while working or through mistakes at the latest.

> *When you fail enough times, you will get it right eventually. You will learn from your mistakes.* [I6, P3]

According to the participants, well-planned and scheduled training regarding the use of the service, as well as equal opportunities and time to participate in these training sessions, are important. The training should also be nonstop so that it would train as many employees as possible and even the shift workers could attend. The participants also pointed out that having *enough knowledge on how to use the service* is important for the success of the implementation. They believed that having enough knowledge would help them achieve confidence in using the service. The participants described problematic situations where the coworkers had taught other coworkers with insufficient knowledge. Furthermore, with such experiences, they mentioned that effective and clear information about the use of the service ensures that it is used correctly.

In addition, 1 of the key practices that the participants pointed out was that the service provider or the organization should *provide different training methods* for employees. One-sided training methods prevented successful implementation. For example, watching training videos alone did not guarantee sufficient skills to successfully use services. However, some of the participants experienced that video training enabled recounting whenever they needed it. In addition to various training methods, participants mentioned that it is important to have, especially after its implementation, written instructions on how to use the service.

The participants hoped that their skills could be assessed to map their training needs. Targeting training according to the level of competence would be useful, as some people may need to learn more basic technical skills, while others already master them well and may be able to cope with shorter training sessions. It was also mentioned that the individuals who embrace the program more easily, such as recent graduates, may find the video training enough and no other forms of training will be needed. The importance of targeting training according to the competence requirements set by the job tasks was also emphasized.

> *I would have wished for targeted training for my own professional group because now they have been general. It would be more efficient if the [training sessions] were more targeted, so you could focus on the things you need in your own work.* [I8, P2]

Reflexive Monitoring: How Do Professionals Reflect on or Appraise the Services' Effects?

The participants noted that after the implementation, it is important that the service be perceived to work well, because experiences of poor usability were believed to jeopardize successful implementation. Thus, the *service should be easy to use* and should not have usability vulnerabilities.

> *Usability and especially the ease of using the service plays a huge role if you want the implementation to be successful.* [I2, P1]

In addition, after the implementation of the service, *monitoring its use* was also considered important. This meant, for example, regular monitoring of the correct use of the system and the use of all included features. The participants were concerned about the misuse of the system due to a lack of sufficient training and monitoring of everyday use.

> *The problem is that no one comes back and asks if you have learned to use the service; a follow-up visit is needed.* [I7, P3]

The participants also pointed out the importance of getting an *opportunity to provide feedback* on the new service after the implementation and that the feedback be used to develop the service. It was essential for the functioning of the feedback channel that it be known by everyone and that the feedback process be perceived as smooth. If it was not perceived as being smooth, there was no desire to give feedback.

> *If you come up with a good idea, taking it forward was not made easy; if it takes a lot of time, it is often left undone.* [I2, P2]

According to the participants, the use of the digital services that are being implemented should be *useful for one's own work.* The service was not used nor recommended, for example, to patients if the participant considered the service unbeneficial for one's own work.

Discussion

Principal Results

This qualitative study identified factors that support successful digital service implementations and should be considered in the future to make sure that the services are integrated into health and social care professionals' routine work. According to professionals' implementation experiences and by using the NPT framework, we suggest 14 practices that should be considered when implementing new services into professionals' clinical work.

To get professionals to understand and make sense of the new service, (1) the communication related to the implementation should be comprehensive and continuous and (2) the implementation process should be consistent. (3) A justification for why the service is going to be implemented should also be given. The best way to engage professionals with the service is (4) to give them opportunities to influence and (5) to make sure that they have a positive attitude toward the upcoming service. To enact the new service into professionals' routine work, it is important that (6) the organization take a supportive approach by providing support from several easy and efficient sources. The professionals should also have (7) enough time to become familiar with the service and have (8) enough know-how about the service. The training should be (9) targeted individually according to skills and work tasks, and (10) it should be diverse. The impact of the implementation on the professionals' work should be evaluated. The service (11) should be easy to use, and (12) usage monitoring should happen. An opportunity (13) to give feedback on the service should also be offered. Moreover, (14) the service should support professionals' work tasks.

Comparison With Prior Work

According to our study, to ensure that the service makes sense (the NPT's component coherence) for the users, the communication related to the implementation should be comprehensive and continuous. In earlier studies, the importance of good information has been highlighted when implementing digital services [14,17,18,37]. However, previous studies do not mention specifically what kind of good communication would best support professionals in their work. In our study, the professionals suggested that the communication should happen from a variety of different channels so that it reaches everyone, including shift workers. Our results highlight that when the implementation process happens under a tight schedule, such as during the COVID-19 pandemic, the importance of comprehensive and continuous communication increases.

In our study, the professionals also highlighted that it is important to justify why the service is going to be implemented in order to ensure that it makes sense for the users. The benefits of using the service seemed to be especially important from the perspective of the professionals' own work. Sanders et al [38] presented in their study that usually the difficulty in implementations is a failure to clarify coherence to the users. If the professionals fail to understand the way of working as helpful and relevant, they may be unwilling to use it. In addition, May et al [37] highlight the importance of coherence; if there is a desire for the service to be normalized into the professionals' work, it is important that the service make sense for the users [27,36]. In the review by Mair et al [19], the sense making was not highlighted as being that important when implementing eHealth systems. However, it is good to note that this review was conducted 10 years ago.

Some earlier studies have pointed out that involvement is 1 of the key contributing factors when implementing new digital services [17-19]. In addition, the NPT [27,36] highlights the importance of engagement and participation (cognitive participation component) with the service to get it normalized into routine work. In our results, the professionals underlined the importance of everyone getting an opportunity to influence and participate in designing the service. Our results therefore support previous results; however, the main problem according to our results is that time for involvement is not given to professionals. The heavy workload and the lack of staffing of health and social care settings have increased in recent years, which may influence the time given for innovation and design [39].

According to our results, to make sure the enactment of the work happens (collective action component), the support should be provided from a variety of fast sources. However, it was interesting that the professionals did not highlight the importance of supervisors' support, whereas previous studies have recognized the lack of support from supervisors as 1 of the major barriers to implementation [17,18,40]. However, we found that the positive attitude toward the service is more important. In addition, enactment of the work happens if people have the necessary skills and training to perform the tasks associated with the services. In our study, it was important for the professionals to have enough knowledge about the service and time for its familiarization. In previous studies, the lack of time [11,18] and the lack of information [18,19] were also experienced as inhibiting factors.

In our study, it was important to use various teaching methods and training, which was targeted according to work tasks.

Professionals also suggested to target the training according to the competence assessment. Previously, the training and competence of professionals have been found to be key factors when implementing new digital services [11,18,19], but competence assessments have received less attention. Mair et al [19] mentioned the importance of training, but they did not specify what kind of training would be required. Therefore, more information is needed in the future about what, how, and how much training should be provided.

In our study, it seems that the evaluation of a new digital service depends on how well the service supports professionals' work tasks and whether users perceive the service use usable (reflexive monitoring). Our results suggest that follow-ups are important, because after the implementation, the users expected someone to monitor whether they were using the services in an appropriate manner. Gagnon et al [17] also found that monitoring the use of the system should happen at the early stages of implementation to ensure immediate response to users' feedback [17]. It is important that the organizations and service providers keep on engaging with the service users after the implementation has taken place. This also provides an opportunity to ensure that the service is used correctly. In addition, in our study, being able to give feedback about the service was considered important because it gave the users the feeling that they can influence. This also gives the organizations and service providers an opportunity to further develop the service based on the users' needs.

The NPT framework seemed to be a suitable choice for our research because our findings were well interpreted with the components of this framework. The NPT offered us a useful tool for organizing the important practices involved in the data, which enabled the development of recommendations for future implementations. We had difficulties sometimes in understanding and applying coding to some of the NPT components, especially coherence and cognitive participation, which are more related to time before implementation. However, even when we used all 4 components, collective action frequently got many mentions, maybe because it describes the enactment of the work most comprehensively. May et al [37] suggested that for future studies, it is important to connect collective action much more closely to the context in implementation studies.

Strengths and Limitations

The strengths of this study include multilevel information about experiences from different professional groups when implementing a different kind of new digital service in the era of the COVID-19 pandemic. The fact that a fairly large number of interviewees (n=30) took part can also be considered a strength. Our results are from the forerunner country in digitalization and from 4 different health centers that are pioneers in digitalization and located in different parts of Finland. Thus, we were able to obtain important information from pioneer organizations for many organizations that are further developing their digital services and systems. A qualitative research method was able to give us a more in-depth overall picture of the situation during the era of COVID-19.

Credibility was established in this qualitative study by following the criteria of reliability presented by Lincoln and Cuba [41]. To establish confidence that the results were true, we collected data generated by 3 interviewers, 1 of whom was a more experienced interviewer. In the Results section, we used quotations to make sure the results were authentic. Credibility and interrater reliability were increased by having several sessions of peer debriefing with the research team about the interviews, the analyses process, and the name of the codes. Dependability was guaranteed by describing the study methods, data collection process, and analysis process as thoroughly as possible. Confirmability was guaranteed by collecting data from different groups of professionals, who were performing different kinds of tasks in a health and social care setting. Transferability was guaranteed by purposely sampling different kinds of professionals to participate in the focus group interviews.

This study had some limitations. One limitation was that because the COVID-19 epidemic got worse during data gathering, we had to conduct some of the interviews using the Microsoft Teams application instead of face-to-face interviews. The focus group interviews' aim is to get people to talk in a group, and with Microsoft Teams, it was more difficult and required more effort from the interviewers. However, fortunately, our interviewers were experienced and were able to plan a strategy for promoting discussion. Interrater reliability was also 1 limitation, because if we had performed the coding by double-coding it, we could have compared the unity of the coders more closely instead of only discussing it [42]. One more of the limitations was that the interviews were conducted in Finnish health centers; compared to other countries, Finland is ahead in digitalization and it can therefore influence implementation attitudes and experiences. So, transferability to other countries must be done with caution, especially related to countries with a low level of digitalization. The interviews were conducted in health centers, where there can be certain types of digital services in use, and therefore the results may not be transferable to other health care contexts. However, a previous review (eg, Mair et al [19]) showed corresponding results even when the environment varied. In addition, given that these 4 health centers were purposely selected because they were advanced in digitalization, it may have also influenced the results. Finally, 1 of the limitations was that most of the professionals who attended the interviews were from the health care sector, and therefore future studies about social care professionals' experiences with digital service implementations should be conducted.

Conclusion

In this study, we examined health and social care professionals' experiences with digital service implementations and identified factors that support successful implementations and should be considered in the future to ensure that the services are integrated into professionals' routine work. Based on the results, we suggested 14 practices for organizations to consider when implementing new digital services. Due to practical reasons, such as limited time and resources and a high number of implementations in organizations, it may not be realistic to expect all the practices to be fully executed. However, these practices can guide organizations to find appropriate ways to

support professionals and help organizations to pursue successful implementations. Our findings can be useful for countries that are beginning their service digitalization or further developing their digital services. For future studies, it is essential to examine implementations in a different phase of the process. The digital services add workload on already busy schedules [17], and thus, it would be beneficial to study how the implementations influence professionals' well-being at work.

Acknowledgments

This study was supported by the Strategic Research Council at the Academy of Finland (project nos. 327145 and 327147) and Finnish Institute for Health and Welfare (THL)-coordinated funding for COVID-19 research included in the Finnish government's supplementary budget. None of these entities had any role in the design of the study or in its data collection, analysis, interpretation of data, or writing of this study. The authors wish to thank all of the interviewees.

Authors' Contributions

JN contributed to conceptualization, methodology, formal analysis, writing (original draft, review and editing), and visualization; A-MK was responsible for conceptualization, methodology, formal analysis, writing (review and editing), and visualization; SK contributed to conceptualization, methodology, formal analysis, and writing (review and editing); EL aided in formal analysis and writing (review and editing); PH, JK, and IK contributed to writing (review and editing); TH aided in conceptualization, writing (review and editing), and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the level of digitalization in studied health centers.
[DOCX File , 26 KB - jmir_v23i12e31668_app1.docx]

Multimedia Appendix 2
Interview guide.
[DOCX File , 26 KB - jmir_v23i12e31668_app2.docx]

References

1. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
2. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
3. 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]
4. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health 2020 Jan 08;6:2055207619898984 [FREE Full text] [doi: 10.1177/2055207619898984] [Medline: 31949918]
5. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital transformation and disruption of the health care sector: internet-based observational study. J Med Internet Res 2018 Mar 27;20(3):e104 [FREE Full text] [doi: 10.2196/jmir.9498] [Medline: 29588274]
6. Abbott PA, Foster J, Marin H, Dykes PC. Complexity and the science of implementation in health IT: knowledge gaps and future visions. Int J Med Inform 2014 Jul;83(7):e12-e22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.10.009] [Medline: 24444700]
7. Glaser J. More on management's role in IT project failures. Healthc Financ Manage 2005 Jan;59(1):82-84. [Medline: 15689017]
8. Heeks R. Health information systems: failure, success and improvisation. Int J Med Inform 2006 Feb;75(2):125-137. [doi: 10.1016/j.ijmedinf.2005.07.024] [Medline: 16112893]
9. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
10. Krause J, Van Lieshout J, Klomp R, Huntink E, Aakhus E, Flottorp S, et al. Identifying determinants of care for tailoring implementation in chronic diseases: an evaluation of different methods. Implement Sci 2014 Aug 12;9(1):102-137 [FREE Full text] [doi: 10.1186/s13012-014-0102-3] [Medline: 25112492]

11. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
12. Ludwick D, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
13. Thygesen H. Technology and Good Dementia care: A Study of Technology and Ethics in Everyday Care Practice. PhD thesis. Norway: Centre for Technology, Innovation and Culture (TIK), University of Oslo; 2009:1-206.
14. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Quarterly 2004 Dec;82(4):581-629. [doi: 10.1111/j.0887-378x.2004.00325.x]
15. 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]
16. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
17. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb 30;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
18. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
19. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 01;90(5):357-364. [doi: 10.2471/blt.11.099424]
20. Bowes A, McColgan G. Implementing telecare for people with dementia: supporting ageing in place in West Lothian, Scotland. J Care Serv Manag 2013 Jul 18;3(3):227-243. [doi: 10.1179/csm.2009.3.3.227]
21. Hibbert D, Mair FS, Angus RM, May C, Boland A, Haycox A, et al. Lessons from the implementation of a home telecare service. J Telemed Telecare 2003 Dec 02;9 Suppl 1(1_suppl):S55-S56. [doi: 10.1258/135763303322196358] [Medline: 12952724]
22. May CR, Finch TL, Cornford J, Exley C, Gately C, Kirk S, et al. Integrating telecare for chronic disease management in the community: what needs to be done? BMC Health Serv Res 2011 May 27;11(1):131 [FREE Full text] [doi: 10.1186/1472-6963-11-131] [Medline: 21619596]
23. Kruse CS, Regier V, Rheinboldt KT. Barriers over time to full implementation of health information exchange in the United States. JMIR Med Inform 2014 Sep 30;2(2):e26 [FREE Full text] [doi: 10.2196/medinform.3625] [Medline: 25600635]
24. Finch T, Mair F, May C. Teledermatology in the UK: lessons in service innovation. Br J Dermatol 2007 Mar;156(3):521-527. [doi: 10.1111/j.1365-2133.2006.07608.x] [Medline: 17300243]
25. Goroll AH, Simon SR, Tripathi M, Ascenzo C, Bates DW. Community-wide implementation of health information technology: the Massachusetts eHealth Collaborative Experience. J Am Med Inform Assoc 2009 Jan 01;16(1):132-139. [doi: 10.1197/jamia.m2899]
26. Foley P, Sutton D, Potter R, Patel S, Gemmel A. International Digital Economy and Society Index 2020. 2021. URL: https://digital-strategy.ec.europa.eu/en/policies/desi [accessed 2010-10-21]
27. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8(oct17 1):63-d6054 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
28. Finch TL, Mair FS, O'Donnell C, Murray E, May CR. From theory to 'measurement' in complex interventions: methodological lessons from the development of an e-health normalisation instrument. BMC Med Res Methodol 2012 May 17;12(1):69 [FREE Full text] [doi: 10.1186/1471-2288-12-69] [Medline: 22594537]
29. 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 7;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Keskimaki I, Tynkkynen LK, Reissell I. Finland: health system review. Health Syst Transit 2019 Aug;21(2):1-166. [doi: 10.1093/oso/9780198860525.003.0010]
31. Nyumba TO, Wilson K, Derrick C, Mukherjee N. The use of focus group discussion methodology: insights from two decades of application in conservation. Methods Ecol Evol 2018 Jan 11;9(1):20-32. [doi: 10.1111/2041-210x.12860]
32. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 May 17;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]

33. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr 16;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
34. Nordmark S, Zingmark K, Lindberg I. Process evaluation of discharge planning implementation in healthcare using normalization process theory. BMC Med Inform Decis Mak 2016 Apr 27;16(48):48-10 [FREE Full text] [doi: 10.1186/s12911-016-0285-4] [Medline: 27121500]
35. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implement Sci 2009 May 21;4(1):29-32 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
36. McEvoy R, Ballini L, Maltoni S, O'Donnell CA, Mair FS, Macfarlane A. A qualitative systematic review of studies using the normalization process theory to research implementation processes. Implement Sci 2014 Jan 02;9(2):2-13 [FREE Full text] [doi: 10.1186/1748-5908-9-2] [Medline: 24383661]
37. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using normalization process theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Jun 07;13(1):80-115 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
38. Sanders T, Foster NE, Ong BN. Perceptions of general practitioners towards the use of a new system for treating back pain: a qualitative interview study. BMC Med 2011 May 09;9(1):49 [FREE Full text] [doi: 10.1186/1741-7015-9-49] [Medline: 21554696]
39. Bee P, Brooks H, Fraser C, Lovell K. Professional perspectives on service user and carer involvement in mental health care planning: a qualitative study. Int J Nurs Stud 2015 Dec 9;52(12):1834-1845 [FREE Full text] [doi: 10.1016/j.ijnurstu.2015.07.008] [Medline: 26253574]
40. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
41. Lincoln YS, Guba EG. Naturalistic Inquiry. Beverly Hills, CA: Sage; 1985.
42. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 22;19(5):160940691989922-160940691989e86. [doi: 10.1177/1609406919899220]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
I-DESI: International Digital Economy and Society Index
NPT: normalization process theory
WHO: World Health Organization

Edited by R Kukafka; submitted 07.07.21; peer-reviewed by B Schreiweis, L Kremer; comments to author 17.08.21; revised version received 11.10.21; accepted 21.10.21; published 01.12.21.

Please cite as:
Nadav J, Kaihlanen AM, Kujala S, Laukka E, Hilama P, Koivisto J, Keskimäki I, Heponiemi T
How to Implement Digital Services in a Way That They Integrate Into Routine Work: Qualitative Interview Study Among Health and Social Care Professionals
J Med Internet Res 2021;23(12):e31668
URL: https://www.jmir.org/2021/12/e31668
doi:10.2196/31668
PMID:34855610

Original Paper

Comparing the Effects of Gamification and Teach-Back Training Methods on Adherence to a Therapeutic Regimen in Patients After Coronary Artery Bypass Graft Surgery: Randomized Clinical Trial

Banafsheh Ghorbani[1], MSc; Alun C Jackson[2,3,4], PhD; Mohammad Noorchenarboo[5], MSc; Mohammad H Mandegar[6], MD; Farshad Sharifi[7], MD, MPH, PhD; Zohrehsadat Mirmoghtadaie[8], PhD; Fatemeh Bahramnezhad[9], PhD

[1]School of Nursing and Midwifery, Tehran University of Medical Sciences, Tehran, Iran

[2]Australian Centre for Heart Health, Melbourne, Australia

[3]Faculty of Health, Deakin University, Geelong, Australia

[4]Centre on Behavioural Health, Hong Kong University, Hong Kong, China

[5]Tehran University of Medical Sciences, Tehran, Iran

[6]Department of Cardiac Surgery, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

[7]Elderly Health Research Center, Endocrinology and Metabolism Population Sciences Research Institute, Tehran University of Medical Sciences, Tehran, Iran

[8]Shaheed Beheshti University of Medical Sciences, Tehran, Iran

[9]School of Nursing & Midwifery, Nursing and Midwifery Care Research Center, Spiritual Health Group, Research Center of Quran, Hadith and Medicine, Tehran University of Medical Sciences, Tehran, Iran

Corresponding Author:
Fatemeh Bahramnezhad, PhD
School of Nursing & Midwifery, Nursing and Midwifery Care Research Center
Spiritual Health Group, Research Center of Quran, Hadith and Medicine
Tehran University of Medical Sciences
Nosrat St.
Tohid Sq.
Tehran, 141973317
Iran
Phone: 98 2166914368
Fax: 98 2161054170
Email: bahramnezhad@sina.tums.ac.ir

Abstract

Background: Patients undergoing coronary artery bypass graft surgery (CABGS) may fail to adhere to their treatment regimen for many reasons. Among these, one of the most important reasons for nonadherence is the inadequate training of such patients or training using inappropriate methods.

Objective: This study aimed to compare the effect of gamification and teach-back training methods on adherence to a therapeutic regimen in patients after CABGS.

Methods: This randomized clinical trial was conducted on 123 patients undergoing CABGS in Tehran, Iran, in 2019. Training was provided to the teach-back group individually. In the gamification group, an app developed for the purpose was installed on each patient's smartphone, with training given via this device. The control group received usual care, or routine training. Adherence to the therapeutic regimen was assessed using a questionnaire on adherence to a therapeutic regimen (physical activity and dietary regimen) and an adherence scale as a pretest and a 1-month posttest.

Results: One-way analysis of variance (ANOVA) for comparing the mean scores of teach-back and gamification training methods showed that the mean normalized scores for the dietary regimen (P<.001, F=71.80), movement regimen (P<.001, F=124.53), and medication regimen (P<.001, F=9.66) before and after intervention were significantly different between the teach-back, gamification, and control groups. In addition, the results of the Dunnett test showed that the teach-back and gamification

groups were significantly different from the control group in all three treatment regimen methods. There was no statistically significant difference in adherence to the therapeutic regimen between the teach-back and control groups.

Conclusions: Based on the results of this study, the use of teach-back and gamification training approaches may be suggested for patients after CABGS to facilitate adherence to the therapeutic regimen.

Trial Registration: Iranian Registry of Clinical Trials IRCT20111203008286N8; https://en.irct.ir/trial/41507

(J Med Internet Res 2021;23(12):e22557) doi:10.2196/22557

KEYWORDS

teach back; gamification; treatment regimen; coronary artery bypass graft; patient training

Introduction

Background

The most important objective of coronary artery bypass graft surgery (CABGS) is to improve the patient's quality of life by reducing angina symptoms and maintaining coronary circulation [1,2]. However, this approach also has complications, despite its many benefits. For example, patients may be exposed to side effects, such as respiratory problems, atelectasis, pneumonia, surgical site infection, gastrointestinal problems, and mood disorders, usually for reasons such as inappropriate adherence to a therapeutic regimen (generally including medication, dietary, and movement regimens) [2,3]. A lack of, or improper, training of patients may be accompanied by serious complications and re-admissions [3,4], although some of these side effects are preventable. Self-care and understanding of one's illness, lifestyle changes, and improvement of the patient's quality of life require the transfer of knowledge and education from the health professional to the patient. Despite recognizing the importance of patient education, several studies have shown that training of these patients may not be effective. Accordingly, effective educational methods should be used for people of different ages and with different levels of literacy to improve the patient's understanding of the disease condition and the treatment process [5,6].

There are several patient-training methods, such as direct (lectures, individual discussion, group discussion, and teach back) and indirect (booklet, pamphlet, CD, animation-based training, and gamification) methods. The choice of the right educational approach is different depending on the level of literacy and educability of the patient, the mastery of the nurse of the relevant method, and the availability of educational facilities, time, and educational space [7-12]. A number of studies have been conducted on the use of direct methods, such as the teach-back method of patient education [13,14]. This is a comprehensive and evidence-based approach that takes into account health literacy and leads to a better understanding of patients and their caregivers from the training provided by giving them information and asking for them to reflect the key points [15,16]. Despite the demonstrated benefits of this method, studies have shown that nurses are either not using this training method or have not found it to be effective [13,17]. Some of the disadvantages of this method are the time-consuming nature of the training, the large amount of content provided, the lack of repetition of the content at different times, and the lack of skill of the trainer [13,18,19]. In addition, some believe that the increasing speed of science and technology has diminished the role of direct teaching methods in the educational process, emphasizing the use of indirect methods, such as animation-, web-, and smartphone-based training.

In recent years, a new method called gamification has been used as a form of indirect patient training. "Gamification" is the term used to define the concept of applying game design and mechanics to nongaming applications. It gives patients the ability to set goals, track progress for achieving them, and get rewarded in return. Gamification also helps to increase users' self-control and is designed to promote positive behavior change. Gamification has been used to train patients with heart failure, myocardial infarction, rheumatoid arthritis, diabetes, breast cancer, and Alzheimer's disease and has also been used for smoking cessation and blood pressure control, yielding positive results [20-22]. However, the method has its critics. Some commentators believe that this training method may lead to addictive behavior, that it is costly, that there is a lack of skilled practitioners, and that there is a lack of appropriate infrastructure for its use [22-24].

Research Aim

Due to the importance of patient education in the development of self-care and the increasing use of modern educational approaches, the aim of this study was to compare the effect of gamification and teach-back training methods on adherence to the therapeutic regimen in patients after CABGS. The primary outcome was a score on adherence to the therapeutic regimen.

Methods

Study Design

This randomized clinical trial was performed with a sample size of 123 people in 2019. The study population consisted of all CABGS patients admitted to the intensive care units of hospitals affiliated to the Tehran University of Medical Sciences, Tehran, Iran. The patients who met the study inclusion criteria were randomly divided into three groups of 41 each using randomized block design and software. To decrease the predictability of allocated groups and ensure randomizing participants in an equal number, we used block randomization with size 4. Allocation in each group was random but equal in size.

Inclusion and Exclusion Criteria

Inclusion criteria were an age range of 18-60 years, an Android phone for the patient, nonuse of psychotropic drugs, the ability to understand and speak Persian, willingness to participate in the study, lack of hearing and speech disorders, and the ability

to receive phone calls after discharge. Exclusion criteria were the patient's unwillingness to continue an education process and acute illness requiring emergency intervention

Intervention

For patients in the gamification group, informed consent was obtained and a pretest questionnaire completed. The training program was installed on each patient's smartphone during discharge, and the researcher explained to the participants how to use the game. The app developed for the program included three main sections (dietary regimen, medication regimen, and movement regimen) and one assessment section at the end of each main section. The training was in the form of animation, images, and sound (Figure 1). In the assessment section, which was in the form of a question, the correct answer received 6 stars (reward, motivation); the screen of the phone was full of stars (excitement), and it moved to the next step. If a false answer was given, 3 stars were deducted from the total score (punishment), and the same step was continued until a correct answer was obtained. For a false answer, a sound (eg, a ding) was made (Textbox 1). Each patient was able to see the sum of their scores and those of other patients on the home screen. The patient was also able to see their score chart relative to those of the other participants, and consistent with social comparison theory, thereby motivating them to learn more in comparison to others. The questionnaire was recompleted 30 days later by an in-person visit to the patients' home.

Figure 1. Sample screenshot from the physical activity section.

Textbox 1. Delban (the app, which means "heart protector") application specifications.

Section 1

Includes the dietary regimen. The scientific content in this section is divided into 7 groups (general recommendations, bread and cereals, meat and beans, dairy, fruits, vegetables, and fat), and finally an assessment is included.

Section 2

Includes the medication regimen. General recommendations are given in three subsections for better patient access. The most commonly prescribed heart medications for these patients include β-blockers (eg, Metoral), antihypertensive drugs (eg, Losartan and Valsartan), antiplatelets (eg, Plavix and aspirin), lipid-lowering drugs (eg, Atorvastatin), diuretics (eg, Lasix, Aldactone, triamterene-H, and hydrochlorothiazide), anticoagulants (eg, warfarin and enoxaparin), and vasodilator drugs (eg, SUSTAC). Finally, an assessment is included.

Section 3

Includes physical activity. General recommendations are given in three parts. Other areas include walking, using a spirometer, breathing activity, sexual activity, returning to work, driving, and foot edema. In addition, the recommended exercises during the second phase of cardiac rehabilitation after surgery are provided to the patient for 4 weeks. The last part includes an assessment.

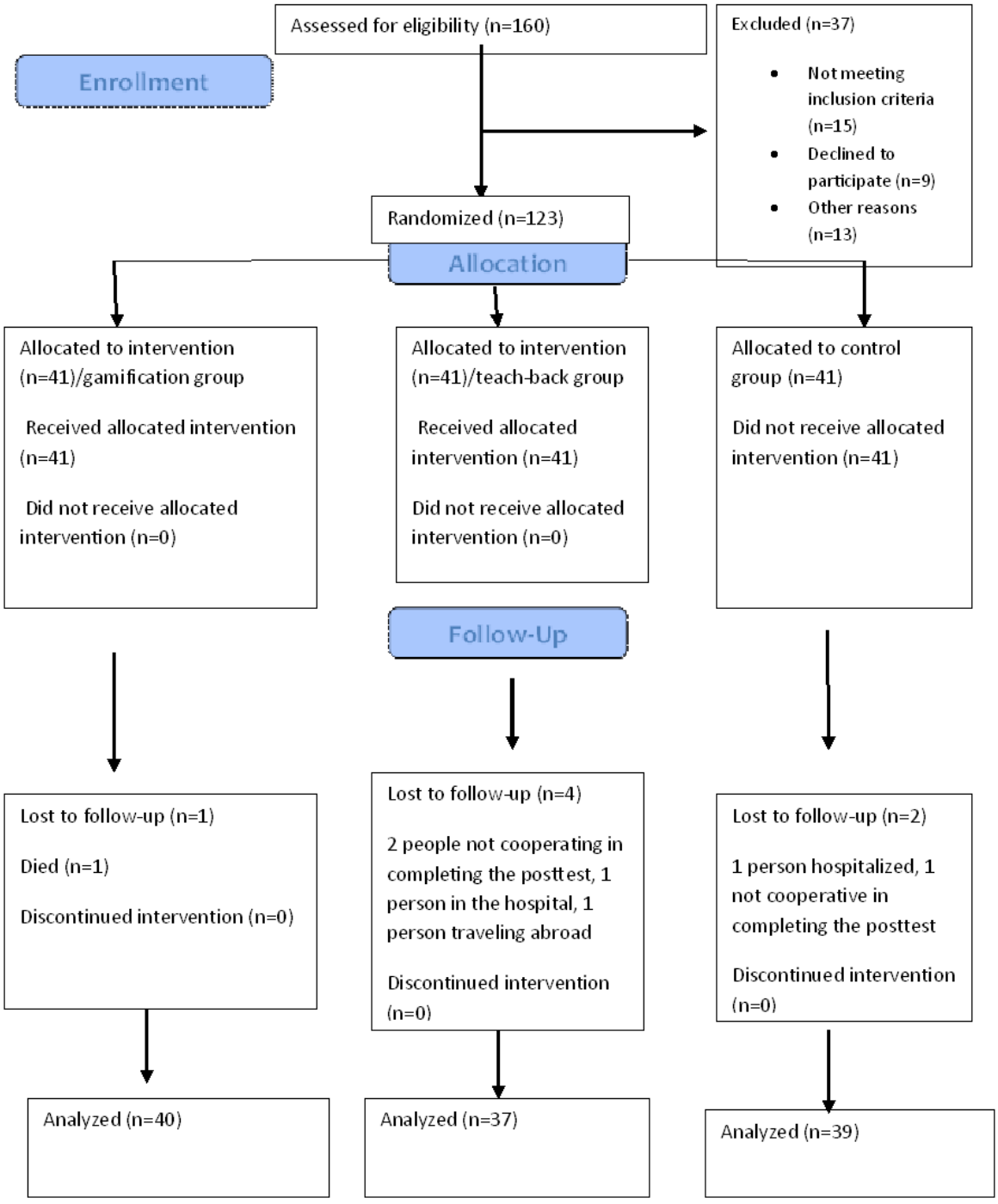
Section 4

Includes scoring and assessment. The patient's assessment section answers the questions considered. The correct answer receives 6 stars (reward, motivation), and the screen of the phone is full of stars (excitement), with progression to the next step. In the case of a false answer, 3 stars are deducted from the total score (punishment), and the same step is continued until the answer is correct. For a false answer, a sound (eg, a ding) is made. The patient sees the sum of their scores and those of other patients on the home screen. The patient also sees their score chart relative to the those of others, which motivates them to learn more in comparison to others.

For patients in the teach-back group, after obtaining informed consent and completing the pretest questionnaire, training was presented in the following steps by the researcher: (1) training information in a simple and understandable language without medical terminology, (2) expressing information in the language of the patient without embarrassment, (3) correcting the patient's misunderstanding, (4) re-asking the patient to make sure the patient was aware of any error, and (5) checking the patient's correct understanding. The duration of training varied from 45 min to a maximum of 60 min, depending on the patient's physical and mental condition. In addition, training was provided individually and outside of the inpatient ward, in a separate room, to prevent data transfer to other patients. In addition, a training booklet was given to patients in the teach-back group in order to access information within a 30-day period. Finally, 30 days later, the questionnaire was recompleted by visiting the patient at home, in person. It should be noted that the educational content of both groups was identical.

In the control group, no training was provided by the researcher, and the group received only the usual training given by the ward nurse. The questionnaire was completed by the patients before discharge and then 30 days later during a home visit (Figure 2).

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of this study.



Data Collection Tools

Prior to training, each patient completed an informed consent form, a demographic questionnaire, a questionnaire on adherence to a therapeutic regimen (physical activity and dietary regimen) designed by Sanaie et al [25], and an adherence scale.

The demographic questionnaire included information about age, sex, education, and source of information about the study. In addition, the patient's health record was referred to obtain details on medical history, past surgery, and type of medication. The questionnaire on adherence to a therapeutic regimen designed by Sanaie et al [25] consists of two sections. The first part contains 30 questions about the patient's dietary regimen (consumption of salt, fat, meat, dairy, etc). Options (4-point Likert scale) are divided into the number of uses per week, with a total score of 100 points per question. Finally, the total adherence rate of the dietary regimen, where 100% represents a total score of 3000, is decided as follows: <50% of the total

score (<1500), undesired adherence; 50%-75% of the total score (1500-2250), relatively desired adherence; and >75% of the total score (>2250), desired adherence.

The second part includes 19 questions about the patient's movement regimen (eg, walking, breathing exercises, and spirometer). The options are scored from never to always (0 to 110) using a 5-point Likert scale. The total adherence rate of the physical activity regimen, where 100% represents a total score of 1900, is considered as follows: <50% of the total score (<950), undesired adherence; 50%-75% of the total score (950-1425), relatively desired adherence; and >75% of the total score (>1425), desired adherence. The reliability of the questionnaire was measured by the Cronbach alpha value (α=.81).

To confirm content validity, the questionnaire was presented to 10 faculty members of the School of Nursing and Midwifery, Tehran University of Medical Sciences, in addition to a nutritionist, a physiotherapist, a sports medicine specialist, a cardiologist, and an interventional cardiologist. After collecting their comments, corrective and suggested comments were applied.

The MMAS has 7 yes/no options (yes=0 and no=1) and 1 5-point Likert scale (never=0, rarely =1, sometimes=2, often =3, almost always=4). A score of 6 and higher is considered to represent the desired level of adherence to the therapeutic regimen. The MMAS has been translated into Persian by the corresponding author and coauthors, and its validity and reliability (Cronbach α=.82) confirmed [26].

Statistical Analysis

Data were analyzed with SPSS Statistics version 20 (IBM) and STATA version 12 (StataCorp) using the Fisher exact test, chi-square test, independent t-test, analysis of variance (ANOVA), and Dunnett test.

Ethical Considerations

This study was approved by the ethics committee of the Tehran University of Medical Sciences, with the code of ethics IR.TUMS.FNM.REC.1398.029, and registered on the database of the Iranian Registry of Clinical Trials (IRCT), with the code IRCT20111203008286N8. Prior to the intervention, the patients signed written informed consent. At the end of the study, the educational content was provided to the control group.

Results

The results of this study showed that 64.68% of the samples in the gamification group, 51.35% in the teach-back group, and 62.66% in the control group were male. Other demographic characteristics of the patients are listed in Table 1.

In addition, the results of this study showed that the mean (SD) scores of dietary and movement regimen adherence were higher in the gamification group than in the other two groups (Table 2).

Table 1. Patients' baseline characteristics.

Demographic characteristics	Control, n (%)	Gamification, n (%)	Teach back, n (%)	Results		
				χ^2	*df*	*P* value
Gender				1.56	2	.45[a]
Female	14 (37.8)	13 (35.1)	18 (48.6)			
Male	23 (62.1)	24 (64.8)	19 (51.3)			
Educational level					8	.17[b]
Primary education	2 (5.4)	5 (13.5)	3 (8.1)			
High school	4 (10.8)	4 (10.8)	2 (5.4)			
Diploma	12 (32.4)	10 (27.03)	5 (13.5)			
Graduate	14 (37.8)	9 (24.3)	21 (56.7)			
Postgraduate	5 (13.5)	9 (24.3)	6 (16.22)			
Income status					4	.88[b]
It is not enough	28 (75.6)	30 (81)	29 (73.3)			
It is enough to some extent	6 (16.2)	5 (13.5)	7 (18.9)			
It is enough	3 (8.1)	2 (5.4)	1 (2.7)			
Chronic condition					3	.80[b]
Diabetes	8 (21.6)	9 (24.3)	7 (18.9)			
Hyperlipidemia	7 (18.9)	6 (16.2)	10 (27)			
Hypertension	21 (56/7)	18 (48.6)	16 (43.2)			
COPD[c]	1 (1.7)	3 (8.1)	2 (5.40)			
Chronic kidney disease	0 (0)	1 (1.7)	2 (5.40)			
Source of information					3	.71[b]
Physician	8 (21.6)	4 (10.8)	2 (5.4)			
Nurse	8 (21.6)	3 (8.1)	1 (2.7)			
Media	3 (8.1)	3 (8.1)	8 (21.6)			
Social network and the internet	18 (48.6)	28 (72.9)	26 (70.2)			
Fat consumption				374/4	2	.12[a]
Low fat	13 (35.1)	7 (18.9)	6 (16.2)			
Ordinary fat	14 (37.8)	27 (67.5)	17 (45.9)			
Fatty	10 (27)	5 (13.5)	14 (37.8)			
Salt consumption				240/0	2	.88[a]
Low salt	13 (35.1)	7 (18.9)	6 (16.21)			
Ordinary salt	14 (37.8)	25 (67.5)	17 (45.9)			
Salty	10 (27.0)	5 (13.5)	14 (37.8)			

[a]Chi-squared test.

[b]The Fisher exact text.

[c]COPD: chronic obstructive pulmonary disease.

Table 2. Comparison of mean (SD) scores of dietary and movement regimen adherence in the study groups.

Study group	Mean (SD)	95% CI[a]
Diet regimen		
Control	–0.7639 (0.5200)	(–0.9885, –0.5393)
Gamification	1.033 (0.655)	(0.815, 1.251)
Teaching	–0.347(0.700)	(–0.575, –0.119)
Movement regimen (physical activity)		
Control	–0.9095 (0.3253)	(–1.0961, –0.7229)
Gamification	1.104 (0.653)	(0.923, 1.285)
Teaching	–0.2745 (0.5270)	(–0.4641, –0.0848)
Medication regimen		
Control	–0.398 (0.679)	(–0.726, –0.071)
Gamification	0.555 (0.924)	(0.238, 0.873)
Teaching	–0.200 (1.107)	(–0.533, 0.133)

[a]CI: confidence interval.

Initially, for each outcome, the analysis was performed without dropout management and then with dropout data considered. The results of the analysis showed that no significant difference was found. The implication of these results is that dropout did not play a role in the significance level of the study results. Moreover, the one-way ANOVA test for comparing the mean scores of teach-back and gamification training methods showed that the mean normalized scores for the dietary regimen (P<.001, F=71.80), movement regimen (P<.001, F=124.53), and medication regimen (P<.001, F=9.66) before and after the intervention were significantly different between teach-back, gamification, and control groups. After the heterogeneity of these scores was determined, the Dunnett test was used to compare teach-back and gamification groups with the control group. In addition, in the dietary and movement regimen adherence, there was no overlap in the gamification method with the teach-back method due to the confidence intervals (CIs) of the mean difference with the control group, which indicates that the gamification approach performs significantly better than the teach-back method (Table 3). In the adherence to the therapeutic regimen, the CIs of the mean difference with the control group overlapped in the gamification approach (0.438, 1.469) and the teach-back approach (-0.330, 0.727), which shows that the gamification approach is not significantly different from the teach-back approach (Figure 3).

Table 3. Dunnett simultaneous tests for level mean–control mean.

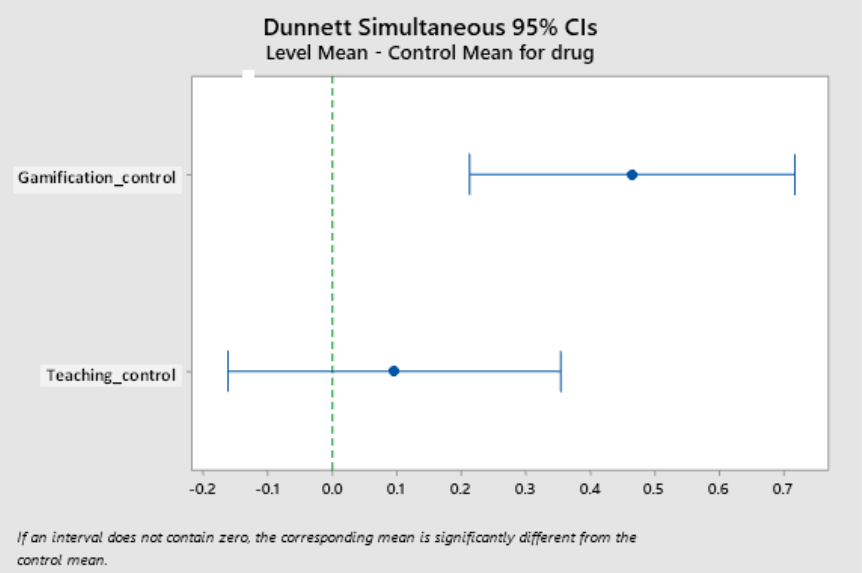
Difference of levels	Mean difference[a]	SE[b] of difference	95% CI[c]	*P* value
Diet regimen				
Gamification-control	1.797	0.157	(1.443, 2.151)	<.001
Teaching-control	0.417	0.161	(0.055, 0.779)	.021
Movement regimen (physical activity)				
Gamification-control	2.013	0.131	(1.719, 2.307)	<.001
Teaching-control	0.635	0.134	(0.334, 0.936)	<.001
Medication regimen				
Gamification-control	0.954	0.230	(0.438, 1.469)	<.001
Teaching-control	0.199	0.235	(-0.330, 0.727)	.61

[a]Individual confidence level=97.29%.

[b]SE: standard error.

[c]CI: confidence interval.

Figure 3. Comparison of CIs between intervention groups (gamification and teach back) and the control group. CI: confidence interval.



Discussion

Principal Findings

This clinical trial was conducted in 2019 with the aim of determining and comparing the effects of gamification and teach-back training methods on adherence to a therapeutic regimen by patients after CABGS. The study found that the gamification training method performs better than the teach-back training method in dietary and movement regimen adherence, but there was no statistically significant difference in adherence to a medication regimen between the two groups. In addition, sensitivity analysis results showed that the dropout did not play a significant role in the significance level of the study results.

A study by Ghanbari et al [27] on the effect of an educational program based on the teach-back method on adherence to a treatment regimen in patients with end-stage renal disease on dialysis showed that the mean score of treatment adherence in 3 areas of hemodialysis, namely drug therapy, fluid restriction, and dietary regimen, in the intervention group in both the posttest (7 days after intervention) and follow-up (30 days after intervention) was significantly higher than in the control group (P<0.01) [27]. In this study, the use of a teach-back training method in relation to movement and dietary regimens showed adherence to the therapeutic regimen, a finding that is in line with Ghanbari et al [27]. However, the use of the teach-back training method in relation to adherence to the medication regimen is inconsistent with their findings. The inconsistency between the results of Ghanbari et al [27] and this study may be due to a longer period of treatment and the periodic nature of hemodialysis compared with open heart surgery and the repetition of teaching during hemodialysis treatment, thus producing a more positive result. In addition, since the medication regimen is such an important pillar of treatment adherence in patients on hemodialysis, more attention may be paid to this area of treatment in patient education using additional information resources, including other patients and treatment staff. Dalir et al's [28] study of the effect of a teach-back intervention aimed at improving self-care in 62 patients with heart failure found that the teach-back training method is effective in improving self-care (P<0.001) [28]. Our study was able to show improvements in relation to movement and diet; however, it did not have the same positive result on the medication regimen that was demonstrated by Dalir et al [28]. The lack of effectiveness in relation to medication adherence in this study may be attributed to the duration of the training. In the Dalir et al [28] study, the patients were taught over 3-4 days, while in this study, the training was presented in a single 45-60-min session. Other studies have also shown the benefits of longer periods of patient education using teach-back training methods. These include White et. al's [29] study of patients with heart failure, whose education occurred over a 13-month period when hospitalized.

Finally, in relation to previous research on gamification effects, Allam et al' [30] RCT on the effect of social support features and gamification on a web-based intervention for patients with rheumatoid arthritis found that physical activity increased in patients in the intervention group who had access to gamification and social support. The use of health care decreased for patients who received social support and for those who received social support and gamification. The study showed that the use of gamification alone or in conjunction with website intervention increases physical activity as well [30]. These significant differences in relation to physical activity using the gamification approach are consistent with our study, as are design elements such as the use of game elements to create excitement and motivation and the awarding of prizes to the "winning" patients.

Limitations

We had to rely on the patients' self-report on exercising at home and following the dietary and medication regimens. In addition, the patients did not record the number of games and browse the application. The app did not include the feature to send reminders and the ability for patients to interact with it. We suggest further studies with a larger sample size and an interactive app with the ability to record usage and to send reminders. In addition, it is recommended that studies be

performed in the presence of active family member inpatient care using these approaches.

Conclusion

Using a game-based smartphone app as a support program to educate patients can affect the patients' adherence to their therapeutic regimen. In addition, with this program, the patients can access the training on their smartphones at any time and place and can repeat the instructions, if necessary. The combination of game elements and patient education together can lead to a better, more engaging learning experience, faster feedback, and more readily available reminders of educational content. However, according to social factor theory, the social signs in multimedia messages (eg, the presentation of an educational agent along with a human voice) causes learners to consider computer-centered learning environments as discourse environments. Signs that speak in the form of a friendly factor on the monitor screen with a human voice and movements increase the ability to transmit positively. The theory of social mediation suggests that bringing verbal (eg, spoken words) and nonverbal (eg, gestures, gaze, and movement) social cues into multimedia environments can simulate human-to-human communication, facilitating engagement of learners in the learning process. According to this theory, by combining a multimedia learning environment and a moving factor as a visual and verbal social symbol, virtual communication between that factor and learners becomes a suitable alternative for human interactions. This study showed that the use of new technology-based approaches can replace previous educational methods. Therefore, by using this method, the process of educating patients can be made more up to date and more attractive by making optimal use of smartphones. Based on the positive results of this study, the teach-back training method is a strategy that can used to increase patients' understanding and is widely accepted by health care organizations as an effective way to communicate information. Easy access, low cost, and the interactive nature of this method are other notable benefits. Training ends when the patient reaches an acceptable level of understanding of the subject. The teach-back approach can be considered an effective and alternative training method instead of traditional ones, such as pamphlets and booklets.

Acknowledgments

This study has been registered with the Nursing and Midwifery Care Research Center (NMCRC), School of Nursing and Midwifery, Tehran University of Medical Sciences (98-01-99-41601) with the code of ethics IR.TUMS.FNM.REC.1398.029. FB expresses his gratitude to this center and the patients participating in this study.

Authors' Contributions

This outcome study was mainly developed by BGH, FBH, ACJ, and FSH. The first draft of the manuscript was written by FB, BGH, ZM, and MN. All the authors confirmed the final draft of the manuscript. The statistical analysis was mainly conducted by FSH and MN. All the authors contributed to the design of the clinical study or the critical revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 94 KB - jmir_v23i12e22557_app1.pdf]

References

1. Sandau K, Boisjolie C, Hodges J. Use of the Minnesota Living With Heart Failure Questionnaire among elderly patients with aortic stenosis: results from a pilot study. J Cardiovasc Nurs 2014;29(2):185-197. [doi: 10.1097/JCN.0b013e318279b76f] [Medline: 23321781]
2. Urden L, Stacy K, Lough M. Critical Care Nursing: Diagnosis and Management, 8th Edition, Kindle Edition, E-book. MO: Mosby; 2018.
3. Plitnik KR. Coronary artery bypass grafting. Nursing2020 Critical Care 2014;9(6):32-37. [doi: 10.1097/01.ccn.0000455850.47502.49]
4. Hasanpour-Dehkordi A, Khaledi-Far A, Khaledi-Far B, Salehi-Tali S. The effect of family training and support on the quality of life and cost of hospital readmissions in congestive heart failure patients in Iran. Appl Nurs Res 2016 Aug;31:165-169. [doi: 10.1016/j.apnr.2016.03.005] [Medline: 27397836]
5. Heydari A, Ziaee ES, Gazrani A. Relationship between awareness of disease and adherence to therapeutic regimen among cardiac patients. Int J Community Based Nurs Midwifery 2015 Jan;3(1):23-30 [FREE Full text] [Medline: 25553331]
6. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2014 Dec 04;21(1):37-44. [doi: 10.1177/1357633x14562734]
7. Darabi L, Shokravi A, Ghaffari M. Comparison of two methods of direct and indirect education on osteoporosis preventive behaviors among female students. Int J Pediatr 2017;5(8):5483-5492. [doi: 10.22038/ijp.2017.20921.1766]

8. Okafor EC, Osuagwu CC, Harman K. The underlying issues in knowledge elicitation. IJIKM 2006;1:095-107. [doi: 10.28945/116]
9. Binder J, Krautz C, Engel K, Grützmann R, Fellner FA, Burger PHM, et al. Leveraging medical imaging for medical education: a cinematic rendering-featured lecture. Ann Anat 2019 Mar;222:159-165. [doi: 10.1016/j.aanat.2018.12.004] [Medline: 30586608]
10. Christou E, Blaney A, Dunn N. Utilising games to generate awareness regarding gender inequality in Parliament. 2015 Presented at: 5th Postgraduate Science and Technology Studies (STS) Conference; March 23-25, 2015; Lancaster University.
11. More SM, Duggal PS, Welkar D, Kolhe S, Patil CD, Chavhan PG. Measurement of course outcomes attainment level for subject by using direct-indirect method. IJIR 2016:1483-1487.
12. Bradshaw M, Hultquist B. Innovative Teaching Strategies in Nursing and Related Health Professions. MA: Jones & Bartlett Learning; 2016.
13. Caplin M, Saunders T. Utilizing teach-back to reinforce patient education: a step-by-step approach. Orthop Nurs 2015;34(6):365-368. [doi: 10.1097/nor.0000000000000197]
14. Talevski J, Wong Shee A, Rasmussen B, Kemp G, Beauchamp A. Teach-back: a systematic review of implementation and impacts. PLoS One 2020;15(4):e0231350 [FREE Full text] [doi: 10.1371/journal.pone.0231350] [Medline: 32287296]
15. Maningat P, Gordon BR, Breslow JL. How do we improve patient compliance and adherence to long-term statin therapy? Curr Atheroscler Rep 2013 Jan 9;15(1):291-223 [FREE Full text] [doi: 10.1007/s11883-012-0291-7] [Medline: 23225173]
16. Miller S, Lattanzio M, Cohen S. "Teach-back" from a patient's perspective. Nursing 2016;46(2):63-64. [doi: 10.1097/01.nurse.0000476249.18503.f5]
17. Kerman SF, Navidian A, Ebrahimi TE, Ghaderi S, Zirak M. Effect of home-based continuous care model on the quality of life of patients with myocardial infarction. Med Surg Nurs J 2017;6(2):17-27 [FREE Full text]
18. Cragun D, Zierhut H. Development of FOCUS-GC: framework for outcomes of clinical communication services in genetic counseling. J Genet Couns 2018 Feb 12;27(1):33-58 [FREE Full text] [doi: 10.1007/s10897-017-0145-0] [Medline: 29022205]
19. Jager AJ, Wynia MK. Who gets a teach-back? Patient-reported incidence of experiencing a teach-back. J Health Commun 2012 Oct;17(Suppl 3):294-302. [doi: 10.1080/10810730.2012.712624] [Medline: 23030577]
20. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
21. Bol N, van Weert JC, de Haes HC, Loos EF, Smets EM. The effect of modality and narration style on recall of online health information: results from a Web-based experiment. J Med Internet Res 2015 Apr 24;17(4):e104 [FREE Full text] [doi: 10.2196/jmir.4164] [Medline: 25910416]
22. Krasulak M. Use of gamification in the process of selection of candidates for the position in the opinion of young adults in Poland. Jagiellonian J Manage 2015;1(3):203-215. [doi: 10.4467/2450114XJJM.15.015.4472]
23. Antonaci A, Klemke R, Stracke C, Specht M. Towards implementing gamification in MOOCs. 2017 Presented at: International Conference on Games and Learning Alliance; 2017; Portugal.
24. Urh M, Vukovic G, Jereb E, Pintar R. The model for introduction of gamification into e-learning in higher education. Procedia: Soc Behav Sci 2015;197:388-397. [doi: 10.1016/j.sbspro.2015.07.154]
25. Sanaie N, Bahramnezhad F, Zolfaghari M, Alhani F. The effect of family-centered empowerment model on treatment plans adherence of patients undergoing coronary artery bypass graft. Crit Care Nurs J 2016 Jul 17;In Press(In Press):23-30. [doi: 10.17795/ccn-6494]
26. Negarandeh R, Mahmoodi H, Noktehdan H, Heshmat R, Shakibazadeh E. Teach back and pictorial image educational strategies on knowledge about diabetes and medication/dietary adherence among low health literate patients with type 2 diabetes. Prim Care Diabetes 2013 Jul;7(2):111-118. [doi: 10.1016/j.pcd.2012.11.001] [Medline: 23195913]
27. Ghanbari R, Nourozi Tabrizi K, Dalvandi A, Kavari H, Noroozi M. Effect of an effect of an educational program based on the teach-back method on adherence to treatment regimen in dialysis patients with end stage renal disease referred to Sina Hospital during 2015-2016. Iran J Rehabilitation Res Nurs 2015;3(4):123-129.
28. Dalir Z, Reihani Z, Mazlom R, Vakilian F. Effect of training based on teach back method on self-care in patients with heart failure. J Maz Univ Med Sci 2016;25(134):209-220 [Persian] [FREE Full text]
29. White M, Garbez R, Carroll M, Brinker E, Howie-Esquivel J. Is "teach-back" associated with knowledge retention and hospital readmission in hospitalized heart failure patients? Journal of Cardiovascular Nursing. J Cardiovasc Nurs 2013;28(2):137-146. [doi: 10.1097/jcn.0b013e31824987bd]
30. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 09;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]

Abbreviations

ANOVA: analysis of variance
CABGS: coronary artery bypass graft surgery

CI: confidence interval
COPD: chronic obstructive pulmonary disease

Edited by G Eysenbach; submitted 16.07.20; peer-reviewed by N Sanaie, J Li; comments to author 28.07.20; revised version received 06.08.20; accepted 05.03.21; published 10.12.21.

Please cite as:
Ghorbani B, Jackson AC, Noorchenarboo M, Mandegar MH, Sharifi F, Mirmoghtadaie Z, Bahramnezhad F
Comparing the Effects of Gamification and Teach-Back Training Methods on Adherence to a Therapeutic Regimen in Patients After Coronary Artery Bypass Graft Surgery: Randomized Clinical Trial
J Med Internet Res 2021;23(12):e22557
URL: https://www.jmir.org/2021/12/e22557
doi:10.2196/22557
PMID:34890346

Original Paper

Active Usage of Mobile Health Applications: Cross-sectional Study

Yang Wang[1], PhD; Tailai Wu[2], PhD; Zhuo Chen[3], PhD

[1]College of Management, Shenzhen University, Shenzhen, China

[2]School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

[3]Department of Health Policy and Management, College of Public Health, University of Georgia, Athens, GA, United States

Corresponding Author:
Tailai Wu, PhD
School of Medicine and Health Management
Huazhong University of Science and Technology
13 Hangkong Road
Qiaokou District
Wuhan, 430030
China
Phone: 86 13477072665
Email: lncle2012@yahoo.com

Abstract

Background: Mobile health applications are being increasingly used for people's health management. The different uses of mobile health applications lead to different health outcomes. Although active usage of mobile health applications is shown to be linked to the effectiveness of mobile health services, the factors that influence people's active usage of mobile health applications are not well studied.

Objective: This paper aims to examine the antecedents of active usage of mobile health applications.

Methods: Grounded on the 3-factor theory, we proposed 10 attributes of mobile health applications that influence the active usage of mobile health applications through consumers' satisfaction and dissatisfaction. We classified these 10 attributes into 3 categories (ie, excitement attributes, performance attributes, and basic attributes). Using the survey method, 494 valid responses were collected and analyzed using structural equation modeling.

Results: Our analysis results revealed that both consumer satisfaction (β=0.351, *t*=6.299, *P*<.001) and dissatisfaction (β=–0.251, *t*=5.119, *P*<.001) significantly influenced active usage. With regard to the effect of attributes, excitement attributes (β=0.525, *t*=12.861, *P*<.001) and performance attributes (β=0.297, *t*=6.508, *P*<.001) positively influenced consumer satisfaction, while performance attributes (β=–0.231, *t*=3.729, *P*<.001) and basic attributes (β=–0.412, *t*=7.132, *P*<.001) negatively influenced consumer dissatisfaction. The results of the analysis confirmed our proposed hypotheses.

Conclusions: Our study provides a novel perspective to study the active usage of mobile health applications. By categorizing the attributes of mobile health applications into 3 categories, the differential effects of different attributes can be tested. Meanwhile, consumer satisfaction and dissatisfaction are confirmed to be independent from each other.

(J Med Internet Res 2021;23(12):e25330) doi:10.2196/25330

KEYWORDS

active usage; mobile health; 3-factor theory; consumer satisfaction; consumer dissatisfaction; medical informatics

Introduction

Background

With the introduction of smartphone devices in the world, mobile health applications are being increasingly used by consumers. By 2019, more than 2.5 billion people owned smartphones worldwide [1], while more than 50% of them had installed mobile health applications in their smartphones by 2017 [2]. The total market size of mobile health applications worldwide was forecasted to reach nearly US $100 billion in 2021 [3]. In the United States, more than 60% of patients use mobile health applications and other digital devices to manage their health [4], while in China, the number of active users of the most popular mobile health applications was over 10.5 million in January 2020 [5]. The most common mobile health applications were related to the weather, fitness, and nutrition in the United States [6] and healthy living information,

measuring/recording of vital signs, and health and medical reminders in China by 2017 [7]. Therefore, mobile health applications have become an important component of individual health management.

Mobile health applications are internet-based applications in mobile devices to support medical and health activities [8]. Generally, mobile health applications provide functions to monitor consumers' health status, record consumers' health-related data, provide medical references, and assist in medical decision making [9]. Mobile health has been shown to affect consumers' health behaviors, including physical activity, diet, alcohol, sexual behavior, and medication adherence [10]. Meanwhile, mobile health has been used to help manage health conditions, including diabetes, asthma, depression, hearing loss, anemia, and migraine [11]. However, the market of mobile health applications is still developing, even with their recognized benefits [12]. In addition to the effect of mobile health applications on people's health status, the usage of mobile health applications affects the success of mobile health services [13].

According to usage status, mobile health application usage could be classified into active and passive. Active usage refers to consumers' high engagement and complete usage of mobile health applications, while passive usage is the occasional and limited use of mobile health applications [14]. Compared with passive usage, active usage of mobile health applications not only uses different functions, such as posting and updating personal information or using nontraditional functions, but also has different antecedents. Since active usage of mobile health applications is linked to the effectiveness of mobile health services, including improving psychological flexibility [15], promoting the maintenance of health-related changes [16], and improving wellness [17], it is meaningful to study the active usage of mobile health applications.

The previous literature has studied the adoption or use of mobile health applications or services based on information system acceptance theories, such as the technology acceptance model (TAM), the theory of reasoned action (TRA), the theory of planned behavior (TPB), and the unified theory of acceptance and use of technology (UTAUT). For example, Hoque and Sorwar [18] extended UTAUT by incorporating technology anxiety and resistance to change in the study of mobile health adoption among older adults and found that only facilitating conditions do not have a significant effect. Alam et al [19] extended UTAUT with perceived reliability and price value and considered gender's moderating role to study mobile health adoption and found that effort expectancy and price value do not influence adoption intention. A comprehensive literature review of mobile health adoption and usage is presented in Multimedia Appendix 1. However, few previous studies have considered the impact of mobile health application attributes and examined the factors that influence the active usage of mobile health applications. Since few prior studies on mobile health applications have examined active usage, we also consider the literature on active use or usage in other contexts. For example, Wu et al [20] studied the active usage of mobile instant messaging applications from an attachment perspective and categorized the predictors into 3 aspects: symbolism, aesthetics, and necessity. Davenport et al [21] found that narcissism is significantly related to active usage in college students but not in adults. A comprehensive literature review of active usage of information systems is presented in Multimedia Appendix 2. Although the previous literature has found several predictors of active usage, none of the predictors reflect the context of mobile health. Considering the importance of active usage of mobile health applications, this study seeks to understand and address the specific research question *What factors influence the active usage of mobile health applications?* Accordingly, we ground our research on the 3-factor theory and propose that consumer satisfaction and dissatisfaction could serve as the mechanisms of the effect of the factors on the active usage of mobile health applications. With regard to the factors, we mainly focus on the attributes of mobile health applications.

By addressing this research question, our study makes 3 important contributions: (1) We explore the factors of active usage of mobile health applications, which has largely been understudied; (2) we categorize and link the attributes to the active usage of mobile health applications based on the 3-factor theory; and (3) we extend and validate the 3-factor theory in the context of mobile health. The rest of the paper is organized as follows: In the next section, we review the literature on the 3-factor theory and consumer satisfaction/dissatisfaction, followed by our model and hypotheses. The subsequent section provides the research methodology and data analysis. Next, the results of the analysis and implications of our study for research and practice are discussed. We conclude with the limitations of this study and avenues for future research in the last section.

Theoretical Background

In this study, the 3-factor theory was used as the overarching theory to construct the research model. Meanwhile, we discuss the theoretical background of satisfaction and dissatisfaction in this section.

3-Factor Theory

The 3-factor theory is the extension of the 2-factor theory, originally proposed to explain job satisfaction in an organizational context [22]. However, the 2-factor theory was criticized for its oversimplification of influential factor categories and no context consideration [23]. Kano et al [24] refined the 2-factor theory and formulated the 3-factor theory to categorize product qualities into 3 factors that meet consumers' needs. The 3 factors are termed as basic factors, excitement factors, and performance factors, aggregating 5 attributes of quality. In Kano's model, basic factors correspond to hygiene factors in the 2-factor theory and represent consumers' basic requirements, which do not affect satisfaction. Excitement factors correspond to motivation factors, which cause excitement but do not cause dissatisfaction when not present. The third category, termed "performance factors," lies between basic and excitement factors, which cause satisfaction when present and dissatisfaction when not present.

The 3-factor theory has been widely applied in different contexts. In the previous literature, the 3-factor theory has been used to categorize online shopping website design attributes [25], test information systems qualities [26], and sort telecom service attributes [27]. Therefore, previous studies have

demonstrated the validity and power of the 3-factor theory in examining the factors influencing usage and adoption and can provide a comprehensive understanding of the antecedents of mobile health application usage.

Consumer Satisfaction and Dissatisfaction

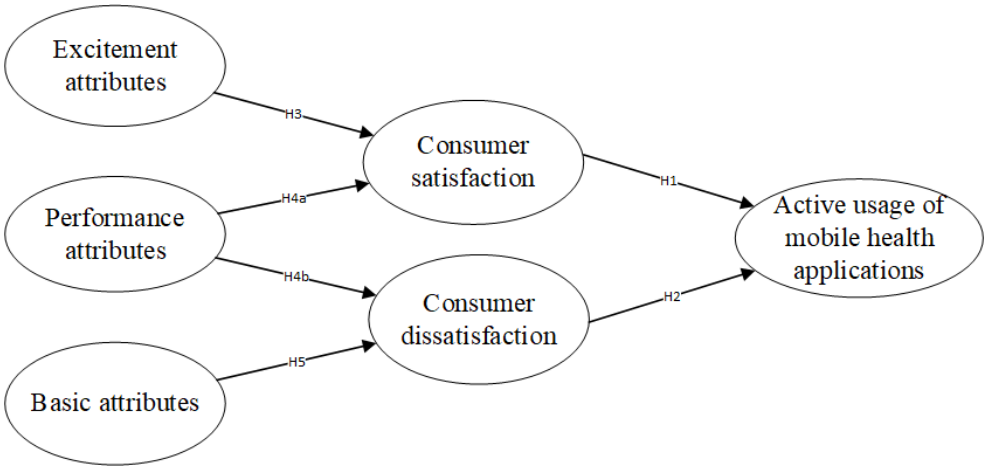
With regard to the relationship between consumer satisfaction and dissatisfaction, previous research has proposed 2 distinct views of the dimensionality of consumer satisfaction [21,28,29]. One view assumes that consumer satisfaction is unidimensional. Thus, consumer satisfaction and dissatisfaction are the 2 ends on a continuum. Another view postulates that consumer satisfaction is bidimensional. Therefore, consumer satisfaction and dissatisfaction are 2 independent constructs. In this study, we argue that consumer satisfaction and dissatisfaction are 2 separate constructs that follow a bidimensional view. The reason consumer satisfaction is not just opposite to consumer dissatisfaction in the current context is that consumer satisfaction and consumer dissatisfaction may also be caused by different factors, which are suggested by the 3-factor theory [23]. Some of the antecedent factors have a significant effect on consumer satisfaction but do not affect consumer dissatisfaction. Meanwhile, consumers' satisfaction levels may not correspond to their dissatisfaction levels. Therefore, consumer satisfaction and dissatisfaction could coexist at the same time [27].

The previous literature also provides evidence that consumer satisfaction is bidimensional. Babin and Griffin [30] demonstrated that the 2-factor model of satisfaction and dissatisfaction could have an acceptable goodness of fit, which indicates that satisfaction and dissatisfaction are distinct. Chen et al [27] found that consumer satisfaction and consumer dissatisfaction have within-construct convergence and between-construct discriminant validities. Kim et al [31] found that factors linked to satisfaction and dissatisfaction of full-service hotels are distinct. Therefore, it is feasible to apply a bidimensional view of consumer satisfaction in our study.

Research Model and Hypothesis Development

According to the 3-factor theory, we proposed that the attributes of mobile health applications will result in satisfaction or dissatisfaction with the applications. Such satisfaction and dissatisfaction may lead to active usage of mobile health applications. Meanwhile, we classified the attributes of mobile health applications into basic attributes, excitement attributes, and performance attributes. The research model is presented in Figure 1.

Figure 1. Research model.



Effect of Consumer Satisfaction and Dissatisfaction

According to the process view, consumer satisfaction refers to the assessment of whether the actual performance of a product or service conforms to one's expectations [32,33]. When the actual performance of a product or service matches or goes beyond consumers' expectations, they feel satisfied. Once they feel satisfied, they improve their attitudes toward the product or service [34,35], engage in positive word-of-mouth activities [36], and increase commitment [37].

Since active usage of mobile health applications can be a reflection of a strong commitment to the relationship between consumers and mobile health applications [38], it could be considered the behavioral representation of consumer engagement in using mobile health applications. Therefore, when consumers feel satisfied using a mobile health application, they will commit and engage in using that mobile health application. Hence, we hypothesize:

> *Hypothesis 1 (H1): Consumer satisfaction positively affects consumers' active usage of mobile health applications.*

Consumer dissatisfaction reflects a consumer's evaluation process during nonconformation of the consumer's expectations and the actual performance of products or services [39]. The previous literature reveals that consumer dissatisfaction might lead to negative behaviors, such as switching [40], complaining [41], or participating in negative word-of-mouth communication [36]. Nonconformity between consumers' expectations and the actual performance of mobile health applications result in negative emotions, such as regret and disappointment [42]. Such negative emotions may only lead to poor commitment of the consumer at best toward using mobile health applications. Hence, we hypothesize:

> *H2: Consumer dissatisfaction negatively affects consumers' active usage of mobile health applications.*

Attributes of Mobile Health Applications

To identify the attributes that influence consumers' active usage of mobile health applications, we conducted interviews as well as a thorough literature review. First, we conducted open-ended interviews with 50 mobile health application users in China by asking them questions such as *Which attributes of mobile health applications influence you to use them actively?* We recorded the responses regarding satisfaction/dissatisfaction in the form of keywords and short sentences. Two PhD students then conducted the content analysis of the recorded responses to identify the items that may reflect the attributes of mobile health applications. Disagreements between them were resolved through discussion. Second, we screened the literature on mobile health adoption to identify possible attributes and established an attribute pool by combining the attributes from the interviews and our thorough literature review. We then selected attributes based on whether they were applicable in mobile health applications. The definitions of the attributes and their sources are summarized in Table 1. Finally, we divided the attributes into 3 categories based on the 3-factor theory.

Table 1. Summary of mobile health applications' attributes.

Number	Attribute	Definition	Source
1	Design aesthetics	The degree of attractiveness or beauty of a mobile health application's interface	Lavie and Tractinsky [43]
2	Customization	The degree to which mobile health application providers tailor their products for different consumers	Srinivasan et al [44]
3	Enjoyment	The degree of fun or pleasure consumers get from using mobile health applications	Van der Heijden [45]
4	Mobility	The degree to which consumers can use mobile health applications regardless of location and time	Hong et al [46]
5	Sociability	The extent to which mobile health applications support social interaction between consumers	Preece [47]
6	Informational support	The degree to which mobile health applications provide information for solving consumers' health concerns	Liang et al [48]
7	Emotional support	The degree to which doctors and other consumers convey their care and understanding to consumers through mobile health applications	Liang et al [48]
8	Perceived security	The extent of security of consumers' private and sensitive information in using mobile health applications	Chang and Chen [49]
9	Technical functionality	Consumers' assessment of mobile health application accessibility, stability, response time, and operation	Ou and Sia [25]
10	Information quality	Consumers' assessment of the accuracy, relevance, and timeliness of information generated by mobile health applications	Delone and Mclean [50]

Combining the definitions of specific attributes with the definitions of the 3 categories, we could categorize the identified attributes into those 3 categories. According to the 3-factor theory, excitement and performance attributes would cause consumer satisfaction, while performance and basic attributes would drive consumer dissatisfaction. With the presence of different attributes, consumer satisfaction and dissatisfaction may coexist. To be specific, we categorized design aesthetics, customization, and enjoyment as excitement attributes because the 3 specific attributes are not the primary concern of consumers but bring additional value to them. Next, we categorized sociability, informational support, and emotional support as performance attributes because these specific attributes can facilitate consumer interaction and communication with health professionals, which are consumers' main purposes of using mobile health applications. Finally, we categorized mobility, perceived security, technical functionality, and information quality as basic attributes because these specific attributes are the basic requirements of any mobile application [51]. This categorization was also preliminarily confirmed by using a card-sorting method [52]. In this study, we treated excitement, performance, and basic attributes as second-order formative constructs that contain the attributes proposed by us and confirmed by the card-sorting method. The relationship between the attributes and consumer satisfaction/dissatisfaction is specified next.

With regard to excitement attributes, design aesthetics, such as the choice of color, shape, or layout of mobile health applications, decide consumers' aesthetics [43]. Meanwhile, design aesthetics can fulfill people's needs for aesthetics in using mobile health applications [53]. Therefore, design aesthetics may influence consumers' satisfaction from using mobile health applications. Customization increases consumers' perceived control toward a mobile health application and signals the quality of the mobile health application [54]. The perceived control and high quality of mobile health applications can influence consumers' satisfaction from using the mobile health applications. With regard to enjoyment, providing consumers with an enjoyable experience is critical for influencing their perception of mobile health applications [45]. Enjoyment in using a mobile health application would be an intrinsic motivation for consumers in using the mobile health application [55]. Therefore, enjoyment reflects the satisfaction of consumers' intrinsic motivation. Hence, we hypothesize:

> *H3: Excitement attributes, which are characterized by design aesthetics, customization, and enjoyment, positively affect consumers' satisfaction.*

With regard to performance attributes, mobile health applications integrate various forms and functions to support the communication and interaction among consumers [47]. Facilitating interaction between health professionals and consumers produces a flow experience in using mobile health applications and increases consumers' satisfaction from using mobile health applications [56]. If mobile health applications cannot facilitate interaction, a consumer may be dissatisfied. Therefore, sociability may lead to satisfaction if mobile health applications have this attribute and dissatisfaction if mobile health applications do not have it. Meanwhile, informational support and emotional support are both dimensions of social support. Health information from mobile health applications could satisfy consumers' needs to deal with their health problems. In contrast, emotional support from doctors and other consumers using mobile health applications might provide emotional support and meet consumers' emotional needs [57]. If mobile health applications cannot afford informational support and emotional support, consumers' health needs may not be satisfied and the evoked negative emotions may make them feel dissatisfied with using mobile health applications. Summarizing the reasonings above, we hypothesize:

H4a: Performance attributes, which are characterized by sociability, informational support, and emotional support, positively affect consumer satisfaction.

H4b: Performance attributes, which are characterized by sociability, informational support, and emotional support, negatively affect consumer dissatisfaction.

With regard to mobility, a basic attribute, which enables consumers to use mobile health applications anywhere and anytime, does not just belong to mobile health applications but also to mobile devices [58]. Given that mobility is shared by other applications on mobile devices [59], it could be considered the basic factor behind using mobile health applications. Therefore, mobility may lead to dissatisfaction if mobile health applications do not have this attribute. Perceived security is 1 of the key drivers for building trust to use any IT [49]. Although some mobile health applications do not integrate mobile shopping functions, and consumers do not suffer any monetary loss, consumers' private information may still be subject to unauthorized access, use, storage, or transmission. Consumers may require application providers to ensure the security of their private information in using mobile health applications. Therefore, perceived security is the basic attribute that may only affect consumers' dissatisfaction. With regard to technical functionality, the technical functions of mobile health applications, such as response speed, stability, and accessibility, may define consumers' usage experience [60], but these functions are not only specific to mobile health applications but also crucial for other IT artifacts. Therefore, technical functionality is a basic attribute and may only influence consumer dissatisfaction if this attribute does not perform well. Regarding information quality, the 3 main aspects that define it are relevance, timeliness, and accuracy of the transmitted information [61]. The quality of communication from using a mobile health application depends upon the quality of information that is transmitted through the mobile health application [62]. Therefore, information quality also serves as the basic condition to complete the main purposes of using mobile health applications and leads to consumer dissatisfaction if it is absent. Hence, we hypothesize:

H5: Basic attributes, which are characterized by mobility, perceived security, technical functionality, and information quality, negatively affect consumer dissatisfaction.

Methods

Measurement Instrument

The survey method was used in this study to validate our proposed research model. The measurement instrument was developed by adapting previously validated scales. The constructs and items sources are presented in Table 2. All items were measured on a 7-point Likert scale with anchors ranging from 1=strongly disagree to 7=strongly agree. In addition, our study included several control variables that measured consumers' characteristics, such as age, gender, education, length, and frequency of using mobile health applications. The length of using mobile health applications reflects how long users have been using mobile health applications, while the frequency of using mobile health applications indicates how many times users use mobile health applications per day.

Table 2. Instrument sources.

Construct	Source
Active usage	Pagani and Mirabello [63]
Satisfaction	Taylor and Baker [64]
Dissatisfaction	Babin and Griffin [30]
Design aesthetics	Lavie and Tractinsky [43]
Customization	Srinivasan et al [44]
Enjoyment	Sun and Zhang [65]
Sociability	Animesh et al [56]
Mobility	Hong et al [46]
Informational support	Liang et al [48]
Emotional support	Liang et al [48]
Information quality	Ou and Sia [25]
Technical functionality	Ou and Sia [25]
perceived security	Cheung and Lee [66]

Since this study was conducted in China and the respondents were primarily Chinese, we translated the survey into Chinese using the back-translation method. The English instrument was first translated into Chinese by 1 bilingual author. Next, another bilingual author back-translated the Chinese version into English. The 2 authors then compared the 2 English versions for inconsistencies. Next, 8 experts of health information systems and 16 users of mobile applications were interviewed to identify ambiguous or repetitive items and suggestions obtained to improve the quality of the survey instrument. Finally, we revised the questionnaire according to the comments and suggestions received. The details of the survey instrument are presented in Multimedia Appendix 3.

Data Collection

The data were collected in China, which is 1 of the largest mobile markets in the world [67]. The online survey was conducted by using a paid service from a popular web survey company in China. The institutional review board of Tongji Medical College, Huazhong University of Science and Technology, China, approved our study procedures (no. 2017S319). Through a 3 weeks' survey in May 2020, we sent the questionnaire to 782 users of Chinese mobile health application users through the online survey company randomly in their users' pool and obtained a total of 674 responses. Therefore, the response rate of our survey was 86.2%.

Given we used an online survey, possible issues, including convenience sample, superrespondents, abnormal respondents, or common method bias, might have existed. By following the guidelines of online surveys, we took several actions that are recommended to ensure data quality [68]. First, our choice of a paid service of online survey company could deal with the issue of a convenience sample since the company has users with diverse backgrounds in different areas. Second, we set screening questions to ask whether the respondents were actual mobile health application users, eliminating respondents who may have produced irrelevant responses. The screening questions included *Do you have mobile health applications in your smartphone or tablet?How many mobile health applications have you installed on your smartphone or tablet?* Actual users could be identified by checking whether they were using mobile health applications according to their answers to these screening questions. Third, to address the issue of superrespondents who were good at filling online questionnaires, we eliminated responses with an unreasonably short time (less than 5 minutes) to finish the questionnaire. Fourth, we eliminated responses where respondents did not correctly answer reverse-coded and attention-trap questions and gave too many same answers for all questions (more than 90%) in order to eliminate haphazard responses. A reverse-coded question is negative worded, and its score needs to be reversed for further data analysis, while an attention-trap question is easy to be answered with only 1 correct option. Finally, we conducted procedural remedies to deal with the common method bias, including randomizing items and using different response formats [57].

After cleaning the collected data, we were left with 494 complete and valid responses. In this sample, most of the respondents were in the age group of 18-30 years, were female (280/494, 56.7%), possessed a college degree, and were familiar with mobile health applications. The demographic distribution of the sample in our study is consistent with the China Internet Network Information Center report of the national profile of mobile internet users [69]. The specific demographic information of our final sample is summarized in Table 3.

Table 3. Demographic information (N=494).

Characteristics	n (%)
Age (years)	
18-25	151 (30.6)
25-30	222 (44.9)
>30	121 (24.5)
Gender	
Male	214 (43.3)
Female	280 (56.7)
Education	
High school	14 (2.8)
College	414 (83.8)
Master degree and above	66 (13.4)
Length of using mobile health applications (years)	
<1	11 (2.3)
1-3	222 (44.9)
>3	261 (52.8)
Frequency of using mobile health applications	
Almost 1 time/day	22 (4.5)
1-10 times/day	280 (56.7)
>10 times/day	192 (38.8)

Data Analysis

This study used partial least squares (PLS), a technique of structural equation modeling, to analyze the data. PLS is a second-generation multivariate causal analysis method and can analyze complex structural equation models [70]. PLS can also be applied in exploratory studies and aims at theory building rather than theory testing. In addition, PLS can model constructs with either formative or reflective indicators. The analysis was conducted using SmartPLS 2.0.3M (SmartPLS GmbH) [71]. Following the 2-stage approach suggested by Anderson and Gerbing [72], we analyzed the measurement model to test reliability and validity, followed by the analysis of a structural model to test our research model. To test reliability and validity, confirmatory factor analysis was conducted, while we implemented a bootstrapping procedure with PLS to analyze the structural model. The results of reliability and validity of our developed measurement instrument and structural model are presented in the Results section.

Results

Reliability and Validity

The indicators of reliability and validity are summarized in Tables 4-6. In Table 4, all Cronbach α and composite reliabilities are >.7, thus demonstrating reliability for all constructs. The value of the average variance extracted (AVE) of each construct was >0.5, thus demonstrating good convergent validity [73]. Based on the results in Tables 5 and 6, each item loading for its assigned construct was >0.7 and even higher for other constructs, while the square roots of the AVEs were all greater than the interconstruct correlations, thus demonstrating discriminant validity [74]. Hence, we concluded that the quality of the measurement model is adequate for testing hypothesized relationships.

Table 4. Construct reliability and convergent validity.

Construct	Composite reliability	AVE[a]	Cronbach α
Active usage	0.84	0.64	.71
Satisfaction	0.92	0.80	.88
Dissatisfaction	0.94	0.84	.91
Design aesthetics	0.93	0.72	.90
Customization	0.95	0.90	.89
Enjoyment	0.92	0.79	.87
Mobility	0.91	0.71	.86
Sociability	0.89	0.74	.82
Information quality	0.89	0.80	.75
Informational support	0.91	0.84	.80
Emotional support	0.91	0.83	.79
Perceived security	0.96	0.92	.91
Technical functionality	0.88	0.72	.80

[a]AVE: average variance extracted.

Table 5. Loading and cross-loading.

Construct	Active usage	Consumer satisfaction	Consumer dissatisfaction	Design aesthetics	Customization	Enjoyment	Mobility	Sociability	Information quality	Informational support	Emotional support	Perceived security	Technical functionality
Active usage 1	*0.82* [a]	0.40	–0.37	0.32	0.13	0.39	0.43	0.41	0.36	0.31	0.30	0.14	0.30
Active usage 2	*0.84* [a]	0.44	–0.39	0.35	0.18	0.43	0.43	0.42	0.37	0.35	0.37	0.25	0.35
Active usage 3	*0.73* [a]	0.38	–0.36	0.43	0.25	0.40	0.35	0.38	0.35	0.30	0.25	0.19	0.31
Consumer satisfaction 1	0.45	*0.89* [a]	–0.55	0.54	0.30	0.59	0.47	0.52	0.49	0.42	0.37	0.36	0.44
Consumer satisfaction 2	0.46	*0.90* [a]	–0.57	0.58	0.35	0.67	0.45	0.55	0.53	0.41	0.33	0.39	0.43
Consumer satisfaction 3	0.45	*0.89* [a]	–0.56	0.57	0.31	0.60	0.39	0.50	0.45	0.38	0.34	0.36	0.42
Consumer dissatisfaction 1	–0.44	–0.56	*0.92* [a]	–0.47	–0.30	–0.48	–0.43	–0.42	–0.44	–0.32	–0.25	–0.35	–0.41
Consumer dissatisfaction 2	–0.44	–0.58	*0.91* [a]	–0.47	–0.31	–0.48	–0.43	–0.47	–0.44	–0.35	–0.34	–0.36	–0.41
Consumer dissatisfaction 3	–0.41	–0.58	*0.92* [a]	–0.50	–0.27	–0.50	–0.42	–0.44	–0.42	–0.37	–0.34	–0.39	–0.42
Design aesthetics 1	0.41	0.52	–0.42	*0.85* [a]	0.43	0.53	0.44	0.46	0.54	0.41	0.31	0.39	0.45
Design aesthetics 2	0.42	0.53	–0.48	*0.85* [a]	0.40	0.57	0.42	0.46	0.52	0.39	0.37	0.42	0.46
Design aesthetics 3	0.33	0.50	–0.37	*0.82* [a]	0.45	0.50	0.30	0.43	0.50	0.33	0.18	0.46	0.36
Design aesthetics 4	0.39	0.53	–0.38	*0.85*	0.46	0.52	0.33	0.45	0.52	0.29	0.20	0.44	0.40
Design aesthetics 5	0.38	0.57	–0.52	*0.86* [a]	0.41	0.62	0.43	0.48	0.54	0.37	0.29	0.46	0.50
Customization 1	0.21	0.31	–0.29	0.47	*0.94* [a]	0.33	0.21	0.32	0.41	0.11	0.03	0.46	0.26
Customization 2	0.23	0.36	–0.31	0.49	*0.96* [a]	0.38	0.25	0.36	0.44	0.16	0.09	0.52	0.30
Enjoyment 1	0.47	0.64	–0.51	0.60	0.34	*0.89* [a]	0.48	0.53	0.53	0.41	0.28	0.45	0.46

Construct	Active usage	Consumer satisfaction	Consumer dissatisfaction	Design aesthetics	Customization	Enjoyment	Mobility	Sociability	Information quality	Informational support	Emotional support	Perceived security	Technical functionality
Enjoyment 2	0.44	0.61	–0.46	0.58	0.32	*0.90* [a]	0.44	0.51	0.48	0.42	0.34	0.39	0.42
Enjoyment 3	0.45	0.60	–0.44	0.55	0.32	*0.88* [a]	0.40	0.46	0.45	0.36	0.29	0.34	0.40
Mobility 1	0.42	0.40	–0.39	0.38	0.20	0.43	*0.85* [a]	0.45	0.43	0.34	0.32	0.20	0.49
Mobility 2	0.41	0.40	–0.37	0.35	0.13	0.41	*0.81* [a]	0.44	0.37	0.44	0.39	0.19	0.41
Mobility 3	0.44	0.39	–0.39	0.43	0.26	0.38	*0.83* [a]	0.48	0.44	0.34	0.29	0.21	0.48
Mobility 4	0.43	0.45	–0.40	0.39	0.22	0.44	*0.87* [a]	0.47	0.47	0.37	0.39	0.19	0.53
Sociability 1	0.46	0.52	–0.44	0.50	0.35	0.52	0.45	*0.86* [a]	0.50	0.36	0.30	0.33	0.44
Sociability 2	0.43	0.51	–0.43	0.48	0.32	0.47	0.48	*0.88* [a]	0.51	0.40	0.33	0.36	0.43
Sociability 3	0.42	0.47	–0.38	0.40	0.26	0.46	0.47	*0.84* [a]	0.45	0.32	0.35	0.30	0.44
Information quality 1	0.40	0.49	–0.43	0.53	0.42	0.48	0.46	0.52	*0.89* [a]	0.36	0.23	0.45	0.50
Information quality 2	0.40	0.50	–0.41	0.58	0.38	0.50	0.45	0.49	*0.89* [a]	0.43	0.31	0.40	0.55
Informational support 1	0.39	0.42	–0.34	0.38	0.11	0.43	0.41	0.40	0.39	*0.91* [a]	0.52	0.24	0.39
Informational support 2	0.35	0.40	–0.36	0.40	0.15	0.38	0.40	0.37	0.41	*0.91* [a]	0.48	0.22	0.41
Emotional support 1	0.34	0.37	–0.31	0.29	0.05	0.33	0.36	0.35	0.27	0.51	*0.92* [a]	0.11	0.40
Emotional support 2	0.36	0.33	–0.30	0.30	0.07	0.30	0.39	0.34	0.28	0.48	*0.90* [a]	0.12	0.36
Perceived security 1	0.21	0.40	–0.37	0.48	0.50	0.43	0.20	0.37	0.45	0.23	0.12	*0.96* [a]	0.33
Perceived security 2	0.26	0.39	–0.39	0.50	0.49	0.42	0.25	0.37	0.47	0.26	0.12	*0.96* [a]	0.33
Technical functionality 1	0.34	0.43	–0.36	0.43	0.27	0.40	0.51	0.45	0.47	0.37	0.35	0.30	*0.82* [a]
Technical functionality 2	0.32	0.36	–0.36	0.42	0.22	0.39	0.48	0.41	0.52	0.38	0.33	0.29	*0.85* [a]
Technical functionality 3	0.36	0.42	–0.43	0.46	0.27	0.43	0.47	0.42	0.51	0.37	0.39	0.29	*0.87* [a]

[a]The square roots of the average variances extracted (AVEs) are in italic.

Table 6. Discriminant validity of constructs.

Construct	Active usage	Consumer satisfaction	Consumer dissatisfaction	Design aesthetics	Customization	Enjoyment	Mobility	Sociability	Information quality	Informational support	Emotional support	Perceived security	Technical functionality
Active usage	*0. 80* [a]	—[b]	—	—	—	—	—	—	—	—	—	—	—
Consumer satisfaction	0.51	*0.90* [a]	—	—	—	—	—	—	—	—	—	—	—
Consumer dissatisfaction	–0.47	–0.62	*0.92* [a]	—	—	—	—	—	—	—	—	—	—
Design aesthetics	0.46	0.63	–0.52	*0.85* [a]	—	—	—	—	—	—	—	—	—
Customization	0.23	0.35	–0.32	0.51	*0.95* [a]	—	—	—	—	—	—	—	—
Enjoyment	0.51	0.70	–0.53	0.65	0.37	*0.89* [a]	—	—	—	—	—	—	—
Mobility	0.51	0.49	–0.46	0.46	0.25	0.50	*0.84* [a]	—	—	—	—	—	—
Sociability	0.51	0.58	–0.48	0.54	0.36	0.56	0.55	*0.86* [a]	—	—	—	—	—
Information quality	0.45	0.55	–0.47	0.62	0.45	0.55	0.51	0.57	*0. 89* [a]	—	—	—	—
Informational support	0.40	0.45	–0.38	0.42	0.14	0.45	0.44	0.42	0.44	*0.91* [a]	—	—	—
Emotional support	0.39	0.39	–0.34	0.32	0.06	0.34	0.41	0.38	0.30	0.55	*0.91* [a]	—	—
Perceived security	0.25	0.41	–0.40	0.52	0.52	0.44	0.24	0.38	0.48	0.25	0.13	*0.96* [a]	—
Technical functionality	0.40	0.48	–0.45	0.51	0.30	0.48	0.57	0.51	0.59	0.44	0.42	0.34	*0.8 5* [a]

[a]The square roots of average variances extracted (AVEs) are in italic.

[b]Not applicable.

We also examined the possibility of common method bias in our study. First, we investigated the correlation coefficients between variables in Table 4 and found that none of the pairs had a high correlation (r>.90) [75]. Second, we conducted the Harman single-factor test using principal component analysis in SPSS Statistics 18.0. In total, 10 factors were extracted, and the first factor in the unrotated solution explained 39.61% of the variation, which is <50% [76]. Third, we employed the marker variable technique to test common method bias [77]. Organizational commitment was chosen as the marker variable since it was theoretically unrelated to our research model. We found organizational commitment had no significant effect on the active usage of mobile health applications (β=.014, P>.05).

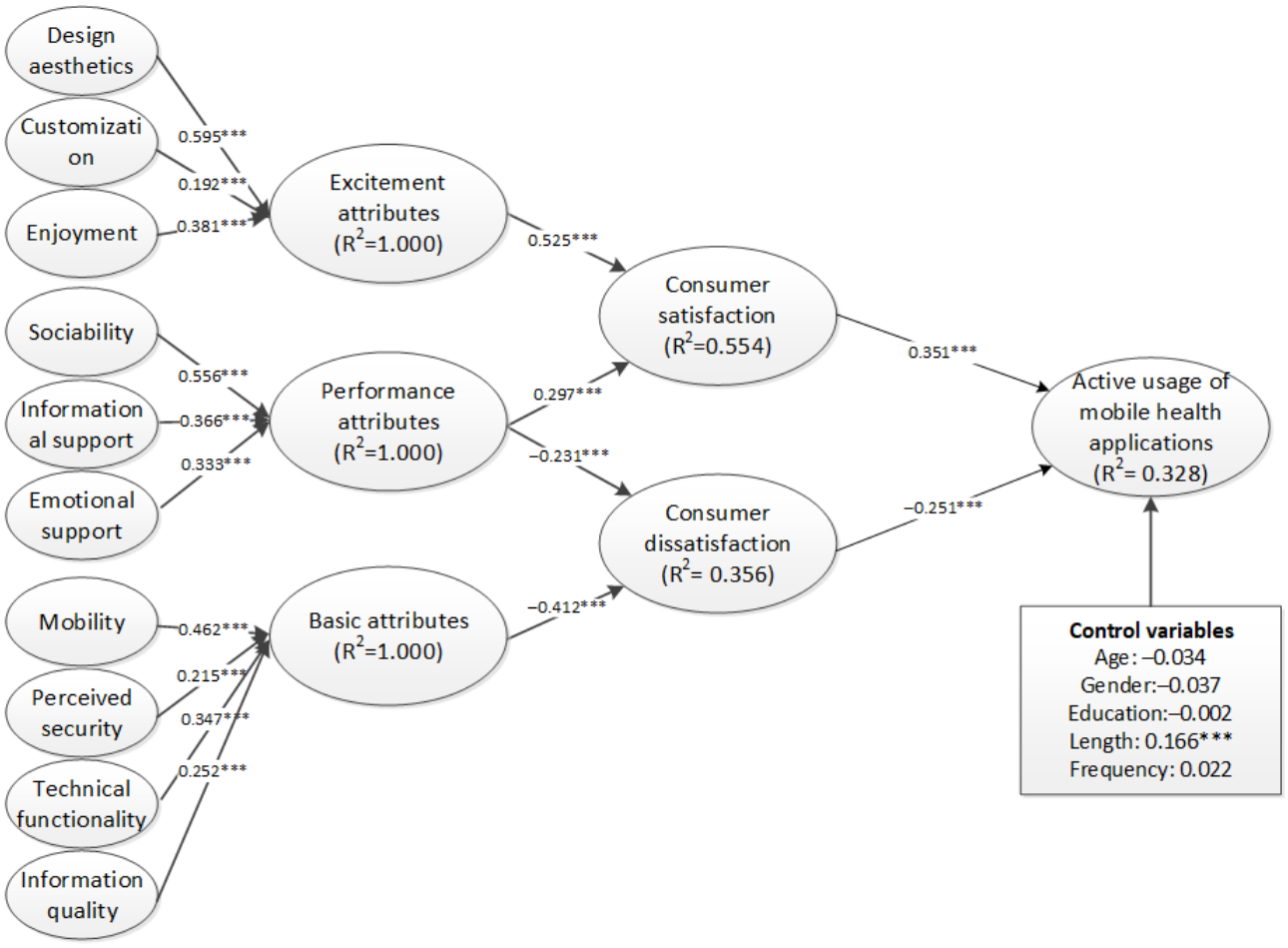
Hypotheses Testing

The descriptive statistics of variables in our research model are presented in Table 7. The results of structural model analysis are summarized in Figure 2. The results reveal that both consumer satisfaction (β=.351, t=6.299, P<.001) and dissatisfaction (β=–.251, t=5.119, P<.001) significantly influence consumers' active usage of mobile health applications, thus supporting H1 and H2. This implies that the 3-factor theory is a useful theoretical perspective for predicting consumers' active usage of mobile health applications.

Table 7. Descriptive statistics of variables.

Construct	Mean (SD)
Active usage	6.00 (0.83)
Satisfaction	5.90 (0.78)
Dissatisfaction	2.07 (0.80)
Design aesthetics	5.70 (0.86)
Customization	4.93 (1.20)
Enjoyment	5.91 (0.77)
Mobility	6.04 (0.84)
Sociability	5.95 (0.82)
Information quality	5.46 (1.06)
Informational support	6.12 (0.72)
Emotional support	6.39 (0.67)
Perceived security	5.21 (1.27)
Technical functionality	5.91 (0.81)

Figure 2. Structural model analysis results.



Note: * Significant at *P*< .05, ** Significant at *P*< .01, *** Significant at *P*< .001

As a second-order formative construct, excitement attributes (β=.525, *t*=12.861, *P*<.001) were revealed to significantly influence consumer satisfaction. Therefore, H3 is supported. Moreover, the corresponding first-order constructs, including design aesthetics, customization, and enjoyment, together comprised excitement factors. With regard to performance attributes, the results also show that they influence both consumer satisfaction (β=.297, *t*=6.508, *P*<.001) and dissatisfaction (β=–.231, *t*=3.729, *P*<.001) significantly. Therefore, H4a and H4b are supported. Meanwhile, the

corresponding first-order constructs, including sociability, informational support, and emotional support, are necessary components of performance factors. With regard to basic attributes (β=–.412, t=7.132, P<.001), which include mobility, perceived security, technical functionality, and information quality, they were found to have a significant influence over consumer dissatisfaction. Therefore, H5 is supported. These results showed that the categorization of basic factors is valid. Finally, the effect of control variables was considered. Only the length of using mobile health applications was significantly associated with active usage of mobile health applications.

Discussion

Principle Findings and Implications

In this study, we identified 10 attributes of mobile health applications that would influence people's active usage. Based on the 3-factor theory, we divided the attributes into 3 categories to correspond to 3 factors, namely excitement attributes, performance attributes, and basic attributes. Excitement and performance attributes are assumed to significantly affect consumer satisfaction, while performance and basic attributes are supposed to significantly impact consumer dissatisfaction. Both consumer satisfaction and dissatisfaction would influence active usage of mobile health applications. Excitement attributes include design aesthetics, customization, and enjoyment; performance attributes include sociability, informational support, and emotional support; and basic attributes include mobility, perceived security, technical functionality, and information quality. To validate the categorization and the effect of attributes on active usage of mobile health applications, we used the survey method and analyzed the data using the PLS technique. The empirical results confirmed our hypothesized differential effects of the 3 attributes on consumer satisfaction and dissatisfaction and then on the active usage of mobile health applications. Meanwhile, all the first-order constructs were also found to be significantly linked to their corresponding second-order constructs. Therefore, the feasibility and validity of the 3-factor theory were manifested in our study. The interestingness and uniqueness of our findings in terms of theoretical and practical implications are further discussed.

From a theoretical perspective, we made several contributions. First, we contributed to health information system usage and adoption literature by studying the active usage of mobile health applications, thus uncovering insights into the use or adoption of mobile health applications. Since most studies do not consider active usage of mobile health applications, as was done in this paper [18,19], considering the importance of active usage toward the effectiveness of mobile health applications, the results of this paper provide a better and deeper understanding of mobile health application adoption and usage.

Second, we also contributed to health information system usage and adoption literature by identifying several influential attributes of active usage of mobile health and categorizing them based on the framework of the 3-factor theory. Previous literature focuses little on the effects of attributes of mobile health applications. This study not only identifies the underlying attributes but also distinguishes the different roles of various attributes of active usage of mobile health applications based on the 3-factor theory. For example, excitement and performance attributes can drive the active usage of mobile health applications, whereas performance and basic attributes could impede the active usage of mobile health applications. Moreover, the study results also reveal a valid categorization of identified attributes.

Third, we contributed to health information system usage and adoption literature by studying the active usage of mobile health applications. Previous studies of active usage of information systems have focused on mobile instant messaging [20], social media [21], or online communities [78-80], a few of them revealing factors driving active usage of mobile health applications. Therefore, our study enriches the literature of active usage of health information systems. At the same time, our study reflects the characteristics of the active usage of mobile health applications by exploring the unique attributes that lead to active usage, such as informational support and emotional support.

Finally, we contributed to the 3-factor theory literature by extending it to the context of mobile health applications. The previous literature mainly applies the 3-factor theory in contexts such as website design or telecom service [25,27], while our study validates the feasibility of the 3-factor theory in the mobile health context. Meanwhile, based on the 3-factor theory, we propose both satisfaction and dissatisfaction as mediators between the antecedents of the usage of mobile health applications and active usage of mobile health applications. Our result provides further support for the distinctiveness between satisfaction and dissatisfaction.

From the practical perspective, this study will help policymakers and medical providers identify active users by using our measurement items of active usage of mobile health applications. For example, the time spent in using mobile health applications, the usage frequency, and some traces of consumers' usage behavior, such as the volume of consumer-generated content in mobile health applications, could be used to compose some indices to locate active users. The thresholds for spending time, usage frequency, generated content, or other tracks could be set to decide who are active users and who are passive users. After identifying the different usage status of different users, policymakers and medical providers could use corresponding strategies to promote the engagement of different users in using mobile health applications.

This study empirically identifies attributes that could be useful for predicting consumers' active usage of mobile health applications and help policymakers and application designers promote proper mobile health applications for consumers. For example, to improve consumers' usage, mobile health applications should focus on excitement attributes and performance attributes, including design aesthetics, customization, enjoyment, sociability, informational support, and emotional support on the premise that mobility, information quality, perceived security, and technical functionality of the mobile health applications are acceptable.

Finally, because consumer satisfaction and dissatisfaction both influence active usage, practitioners should not only emphasize

consumer satisfaction in using mobile health applications but also reduce consumer dissatisfaction from using mobile health applications. Once a consumer feels dissatisfied with using any aspect of a mobile health application, they may become a passive user from being an active user. If they continue to remain dissatisfied with using the mobile health application, and the barriers to switching are low, then they may switch to other mobile health applications.

Limitations and Future Research

The results of this study could be interpreted in light of its limitations. First, although we identified several antecedents of mobile health application usage, the explained variances of active usage of mobile health applications still have the potential to be improved. Our literature review of mobile health application usage and adoption revealed that social, personal, and motivational factors could influence mobile health application usage and adoption, whose effects on active usage of mobile health applications are worth to be examined. Future studies may examine other factors to better understand consumers' active usage.

Second, the generalizability may be restricted because our sample was restricted to Chinese consumers. In China, the most popular mobile health application is Ping An Good Doctor [5], but in other countries, other mobile health applications, such as Pedometer in Sweden [78] or Samsung Health in the United States [79], are leading. There are differences between mobile health applications in different countries. Future studies may conduct cross-country comparisons to better generalize the results of this study.

Finally, our study was mainly a cross-sectional one, where constructs are measured at the same point of time. However, since consumer behavior and mobile health applications are both dynamic, the results may change over time. Therefore, the cross-sectional method may not reflect the dynamics of mobile health application usage. A longitudinal study with a multimethod approach may help address this issue.

Conclusion

In this study, we explored the factors influencing consumers' active usage of mobile health applications based on the 3-factor theory. According to the 3-factor theory, we focused on the attributes of mobile health applications and divided them into 3 categories: excitement, performance, and basic. The 3 categories of attributes are assumed to influence the active usage of mobile health applications through consumer satisfaction and dissatisfaction. The proposed relationships were validated by using a survey method. The analysis results not only imply that consumer satisfaction and dissatisfaction are independent from each other, but also confirm the categorization of attributes for active usage of mobile health applications. Meanwhile, our study could inspire designers and policymakers to make patients actively use mobile health applications.

Acknowledgments

This work was partially supported by grants from the National Natural Science Foundation of China (grants 71801100, 72004072, and 72004071), the Humanities and Social Science Foundation of Ministry of Education of China (grants 20YJC870016 and 20YJCZH169), and the Fundamental Research Funds for the Central Universities (grant 2019kfyXJJS172).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature review of mobile health adoption and usage.
[DOCX File , 25 KB - jmir_v23i12e25330_app1.docx]

Multimedia Appendix 2
Literature review of active usage.
[DOCX File , 23 KB - jmir_v23i12e25330_app2.docx]

Multimedia Appendix 3
Survey instruments.
[DOCX File , 22 KB - jmir_v23i12e25330_app3.docx]

References

1. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-04-16]
2. Hammer R. Mobile Healthcare Technology Statistics. 2017. URL: https://getreferralmd.com/2015/08/mobile-healthcare-technology-statistics/ [accessed 2020-04-16]
3. Mikulic M. Total mhealth Market Size Forecast Worldwide 2016-2025. 2018. URL: https://www.statista.com/statistics/938544/mhealth-market-size-forecast-globally/ [accessed 2020-04-16]

4. McCarthy J. Survey: 64 percent of Patients Use a Digital Device to Manage Health. 2017. URL: https://www.mobihealthnews.com/content/survey-64-percent-patients-use-digital-device-manage-health [accessed 2020-04-17]
5. Thomala LL. Leading Medical Apps by Monthly Active Users in China 2020. URL: https://www.statista.com/statistics/1043739/china-leading-healthcare-apps-monthly-active-users/ [accessed 2020-04-17]
6. Kunst A. Share of Categories of Mobile Health Apps Used among U.S. Consumers 2017. 2019. URL: https://www.statista.com/statistics/378850/top-mobile-health-application-categories-used-by-us-consumers/ [accessed 2020-04-17]
7. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
8. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2020-04-17]
9. Varshney U. Mobile health: four emerging themes of research. Decis Support Syst 2014 Oct;66:20-35. [doi: 10.1016/j.dss.2014.06.001]
10. 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]
11. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
12. Yu X, Han H, Du J, Wei L, Li C, Zhang X. mHealth in China and the United States: How Mobile Technology Is Transforming Healthcare in the World's Two Largest Economies. 2014. URL: https://www.brookings.edu/research/mhealth-in-china-and-the-united-states-how-mobile-technology-is-transforming-health-care-in-the-worlds-two-largest-economies/ [accessed 2020-04-17]
13. Gelderman M. The relation between user satisfaction, usage of information systems and performance. Inf Manag 1998 Aug;34(1):11-18. [doi: 10.1016/s0378-7206(98)00044-5]
14. Kang J, Tang L, Fiore AM. Enhancing consumer–brand relationships on restaurant Facebook fan pages: maximizing consumer benefits and increasing active participation. Int J Hosp Manag 2014 Jan;36:145-155. [doi: 10.1016/j.ijhm.2013.08.015]
15. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
16. Mattila E, Orsama A, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 29;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]
17. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013 Jul 10;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
18. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
19. Alam MZ, Hoque MR, Hu W, Barua Z. Factors influencing the adoption of mHealth services in a developing country: a patient-centric study. Int J Inf Manag 2020 Feb;50:128-143. [doi: 10.1016/j.ijinfomgt.2019.04.016]
20. Wu T, Lu Y, Gong X, Gupta S. A study of active usage of mobile instant messaging application: an attachment theory perspective. Inf Dev 2016 Jul 09;33(2):153-168. [doi: 10.1177/0266666916646814]
21. Davenport SW, Bergman SM, Bergman JZ, Fearrington ME. Twitter versus Facebook: exploring the role of narcissism in the motives and usage of different social media platforms. Comput Hum Behav 2014 Mar;32:212-220. [doi: 10.1016/j.chb.2013.12.011]
22. Maddox RN. Two-factor theory and consumer satisfaction: replication and extension. J Consum Res 1981 Jun;8(1):97. [doi: 10.1086/208845]
23. Matzler K, Renzl B. Assessing asymmetric effects in the formation of employee satisfaction. Tour Manag 2007 Aug;28(4):1093-1103. [doi: 10.1016/j.tourman.2006.07.009]
24. Kano N, Seraku N, Takahashi F, Tsuji S. Attractive quality and must-be quality. JSQC 1984;14(2):147-156.
25. Ou CX, Sia CL. Consumer trust and distrust: an issue of website design. Int J Hum-Comput Stud 2010 Dec;68(12):913-934. [doi: 10.1016/j.ijhcs.2010.08.003]
26. Pak H, Park S, Jang I, Choe Y. A measurement of non-linear relationship between IS qualities and user satisfaction. 2010 Aug 12 Presented at: The Sixteenth Americas Conference on Information Systems; August 2010; Lima, Peru p. 12-15.
27. Chen A, Lu Y, Gupta S, Xiaolin Q. Can customer satisfaction and dissatisfaction coexist? An issue of telecommunication service in China. J Inf Technol 2014 Sep 01;29(3):237-252. [doi: 10.1057/jit.2013.26]
28. Wan J, Zhao L, Lu Y, Gupta S. Evaluating app bundling strategy for selling mobile apps: an ambivalent perspective. ITP 2017 Mar 06;30(1):2-23. [doi: 10.1108/itp-08-2015-0210]

29. Westbrook RA, Oliver RL. The dimensionality of consumption emotion patterns and consumer satisfaction. J Consum Res 1991 Jun;18(1):84. [doi: 10.1086/209243]
30. Babin BJ, Griffin M. The nature of satisfaction: an updated examination and analysis. J Bus Res 1998 Feb;41(2):127-136. [doi: 10.1016/s0148-2963(97)00001-5]
31. Kim B, Kim S, Heo CY. Analysis of satisfiers and dissatisfiers in online hotel reviews on social media. Int J Contemp Hosp Manag 2016 Sep 12;28(9):1915-1936. [doi: 10.1108/ijchm-04-2015-0177]
32. Zhou Z, Li X. Enablers and inhibitors of discontinuous use in social networking sites: a study on Weibo. 2018 Presented at: PACIS 2018; June 26–30, 2018; Yokohama, Japan.
33. Yi Y. A critical review of consumer satisfaction. Rev Mark 1990;4(1):68-123.
34. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 1980 Nov;17(4):460 [FREE Full text] [doi: 10.2307/3150499]
35. Oliver RL, DeSarbo WS. Response determinants in satisfaction judgments. J Consum Res 1988 Mar;14(4):495-507. [doi: 10.1086/209131]
36. Anderson EW. Customer satisfaction and word of mouth. J Serv Res 2016 Jun 29;1(1):5-17. [doi: 10.1177/109467059800100102]
37. Beatson A, Coote LV, Rudd JM. Determining consumer satisfaction and commitment through self-service technology and personal service usage. J Mark Manag 2006 Aug;22(7-8):853-882. [doi: 10.1362/026725706778612121]
38. Mollen A, Wilson H. Engagement, telepresence and interactivity in online consumer experience: reconciling scholastic and managerial perspectives. J Bus Res 2010 Sep;63(9-10):919-925. [doi: 10.1016/j.jbusres.2009.05.014]
39. Singh J. A typology of consumer dissatisfaction response style. J Retail 1990;66(1):57-99.
40. Richins ML. A multivariate analysis of responses to dissatisfaction. JAMS 1987 Sep;15(3):24-31. [doi: 10.1007/bf02722168]
41. Singh J. Consumer complaint intentions and behavior: definitional and taxonomical issues. J Mark 2018 Nov 02;52(1):93-107. [doi: 10.1177/002224298805200108]
42. Smith CA, Ellsworth PC. Patterns of cognitive appraisal in emotion. J Pers Soc Psychol 1985 Apr;48(4):813-838. [Medline: 3886875]
43. Lavie T, Tractinsky N. Assessing dimensions of perceived visual aesthetics of web sites. Int J Hum-Comput Stud 2004 Mar;60(3):269-298. [doi: 10.1016/j.ijhcs.2003.09.002]
44. Srinivasan SS, Anderson R, Ponnavolu K. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. J Retail 2002 Mar;78(1):41-50. [doi: 10.1016/S0022-4359(01)00065-3]
45. Van der Heijden H. User acceptance of hedonic information systems. MIS Quarterly 2004;28(4):695-704. [doi: 10.2307/25148660]
46. Hong S, Thong JYL, Moon J, Tam K. Understanding the behavior of mobile data services consumers. Inf Syst Front 2008 Jun 12;10(4):431-445. [doi: 10.1007/s10796-008-9096-1]
47. Preece J. Sociability and usability in online communities: determining and measuring success. Behav Inf Technol 2010 Nov 08;20(5):347-356. [doi: 10.1080/01449290110084683]
48. Liang T, Ho Y, Li Y, Turban E. What drives social commerce: the role of social support and relationship quality. Int J Electron Commer 2014 Dec 10;16(2):69-90. [doi: 10.2753/jec1086-4415160204]
49. Chang HH, Chen SW. Consumer perception of interface quality, security, and loyalty in electronic commerce. Inf Manag 2009 Oct;46(7):411-417. [doi: 10.1016/j.im.2009.08.002]
50. DeLone WH, McLean ER. Int J Electron Commer 2014 Dec 08;9(1):31-47. [doi: 10.1080/10864415.2004.11044317]
51. Matzler K, Fuchs M, Schubert A. Employee satisfaction: does Kano's model apply? Total Qual Manag Bus 2010 Aug 25;15(9-10):1179-1198. [doi: 10.1080/1478336042000255569]
52. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
53. Sanchez-Franco MJ, Rondan-Cataluña FJ. Virtual travel communities and customer loyalty: customer purchase involvement and web site design. Electron Commer Res Appl 2010 Mar;9(2):171-182. [doi: 10.1016/j.elerap.2009.05.004]
54. Dellaert BGC, Dabholkar PA. Increasing the attractiveness of mass customization: the role of complementary on-line services and range of options. Int J Electron Commer 2014 Dec 08;13(3):43-70. [doi: 10.2753/jec1086-4415130302]
55. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
56. Animesh A, Pinsonneault A, Yang SB, Oh W. An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. MIS Quarterly 2011;35(3):789-810. [doi: 10.2307/23042809]
57. Wu T, Deng Z, Feng Z, Gaskin DJ, Zhang D, Wang R. The effect of doctor-consumer interaction on social media on consumers' health behaviors: cross-sectional study. J Med Internet Res 2018 Feb 28;20(2):e73 [FREE Full text] [doi: 10.2196/jmir.9003] [Medline: 29490892]
58. Perry M, O'hara K, Sellen A, Brown B, Harper R. Dealing with mobility: understanding access anytime, anywhere. ACM Trans Comput-Hum Interact 2001 Dec;8(4):323-347. [doi: 10.1145/504704.504707]
59. Lynn M. Scarcity effects on value: a quantitative review of the commodity theory literature. Psychol Mark 1991;8(1):43-57. [doi: 10.1002/mar.4220080105]

60. Kim J, Lee J, Han K, Lee M. Businesses as buildings: metrics for the architectural quality of internet businesses. Inf Syst Res 2002 Sep;13(3):239-254. [doi: 10.1287/isre.13.3.239.79]
61. Seddon PB. A respecification and extension of the DeLone and McLean model of IS success. Inf Syst Res 1997 Sep;8(3):240-253. [doi: 10.1287/isre.8.3.240]
62. Chae M, Kim J, Kim H, Ryu H. Information quality for mobile internet services: a theoretical model with empirical validation. Electron Mark 2002 Jan 1;12(1):38-46. [doi: 10.1080/101967802753433254]
63. Pagani M, Mirabello A. The influence of personal and social-interactive engagement in social TV web sites. Int J Electron Commer 2014 Dec 10;16(2):41-68. [doi: 10.2753/jec1086-4415160203]
64. Taylor SA, Baker TL. An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions. J Retail 1994 Jun;70(2):163-178. [doi: 10.1016/0022-4359(94)90013-2]
65. Sun H, Zhang P. Causal relationships between perceived enjoyment and perceived ease of use: an alternative approach. JAIS 2006 Sep;7(9):618-645. [doi: 10.17705/1jais.00100]
66. Cheung CM, Lee MK. Understanding consumer trust in internet shopping: a multidisciplinary approach. J Am Soc Inf Sci 2006 Feb 15;57(4):479-492. [doi: 10.1002/asi.20312]
67. Thomala L. Mobile Apps in China: Statistics & Facts. 2020. URL: https://www.statista.com/topics/5577/mobile-apps-in-china/ [accessed 2020-04-17]
68. Lowry PB, D'Arcy J, Hammer B, Moody GD. "Cargo Cult" science in traditional organization and information systems survey research: a case for using nontraditional methods of data collection, including Mechanical Turk and online panels. J Strateg Inf Syst 2016 Oct;25(3):232-240. [doi: 10.1016/j.jsis.2016.06.002]
69. China Internet Network Information Center. 44th Statistical Report on Internet Development in China. 2019. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/201911/P020191112539794960687.pdf [accessed 2020-04-17]
70. Urbach N, Ahlemann F. Structural equation modeling in information systems research using partial least squares. J Inf Technol Theor Appl 2010;11(2):5-40.
71. Ringle C, Wende S, Will A. SmartPLS 2.0.M. Hamburg: SmartPLS; 2005. URL: https://www.smartpls.com/smartpls2/download [accessed 2020-04-17]
72. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
73. Nunnally J. Psychometric Theory. New York: McGraw-Hill; 1967.
74. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
75. Bagozzi RP, Yi Y, Phillips LW. Assessing construct validity in organizational research. Adm Sci Quarterly 1991 Sep;36(3):421. [doi: 10.2307/2393203]
76. Podsakoff PM, Organ DW. Self-reports in organizational research: problems and prospects. J Manag 2016 Sep 04;12(4):531-544. [doi: 10.1177/014920638601200408]
77. Lindell MK, Whitney DJ. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121. [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
78. Stoll J. Leading Health and Fitness Apps in the Google Play Store in Sweden in October 2019, by Number of Downloads. 2019. URL: https://www.statista.com/statistics/691338/leading-mobile-health-and-fitness-apps-google-play-downloads-sweden/ [accessed 2020-04-17]
79. Clement J. Leading Android Health Apps in the U.S. 2020, by Downloads. 2020. URL: https://www.statista.com/statistics/699054/leading-google-play-health-usa-downloads/ [accessed 2020-04-17]
80. Khansa L, Ma X, Liginlal D, Kim SS. Understanding members' active participation in online question-and-answer communities: a theory and empirical analysis. J Manag Inf Syst 2015 Aug 28;32(2):162-203. [doi: 10.1080/07421222.2015.1063293]

Abbreviations

AVE: average variance extracted
PLS: partial least squares
TAM: technology acceptance model
TPB: theory of planned behavior
TRA: theory of reasoned action
UTAUT: unified theory of acceptance and use of technology

Edited by A Mavragani; submitted 28.10.20; peer-reviewed by X Zhang, M Zuo, J Zhou, Z Xie, Q Xiao, J Zhou; comments to author 10.12.20; revised version received 13.01.21; accepted 27.11.21; published 22.12.21.

Please cite as:
Wang Y, Wu T, Chen Z
Active Usage of Mobile Health Applications: Cross-sectional Study
J Med Internet Res 2021;23(12):e25330
URL: https://www.jmir.org/2021/12/e25330
doi:10.2196/25330
PMID:34941545

Original Paper

The Effects of Virtual Reality Treatment on Prefrontal Cortex Activity in Patients With Social Anxiety Disorder: Participatory and Interactive Virtual Reality Treatment Study

Hojun Lee[1,2], MD, PhD; JongKwan Choi[3], PhD; Dooyoung Jung[4], MD, PhD; Ji-Won Hur[5], PhD; Chul-Hyun Cho[6,7], MD, PhD

[1]Department of Psychiatry, School of Medicine, Keimyung University, Daegu, Republic of Korea

[2]Department of Psychiatry, Keimyung University Dongsan Medical Center, Daegu, Republic of Korea

[3]OBELAB Inc, Seoul, Republic of Korea

[4]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

[5]School of Psychology, Korea University, Seoul, Republic of Korea

[6]Department of Psychiatry, College of Medicine, Chungnam National University, Daejeon, Republic of Korea

[7]Department of Psychiatry, Chungnam National University Sejong Hospital, Sejong, Republic of Korea

Corresponding Author:
Chul-Hyun Cho, MD, PhD
Department of Psychiatry
College of Medicine
Chungnam National University
82 Munhwa-ro
Jung-gu
Daejeon, 35015
Republic of Korea
Phone: 82 044 995 4775
Email: david0203@gmail.com

Abstract

Background: Attempts to use virtual reality (VR) as a treatment for various psychiatric disorders have been made recently, and many researchers have identified the effects of VR in psychiatric disorders. Studies have reported that VR therapy is effective in social anxiety disorder (SAD). However, there is no prior study on the neural correlates of VR therapy in patients with SAD.

Objective: The aim of this study is to find the neural correlates of VR therapy by evaluating the treatment effectiveness of VR in patients with SAD using portable functional near-infrared spectroscopy (fNIRS).

Methods: Patients with SAD (n=28) were provided with 6 sessions of VR treatment that was developed for exposure to social situations with a recording system of each participant's self-introduction in VR. After each VR treatment session, the first-person view (video 1) and third-person view (video 2) clips of the participant's self-introduction were automatically generated. The functional activities of prefrontal regions were measured by fNIRS while watching videos 1 and 2 with a cognitive task, before and after whole VR treatment sessions, and after the first session of VR treatment. We compared the data of fNIRS between patients with SAD and healthy controls (HCs; n=27).

Results: We found that reduction in activities of the right frontopolar prefrontal cortex (FPPFC) in HCs was greater than in the SAD group at baseline (t=–2.01, P=.049). Comparing the frontal cortex activation before and after VR treatment sessions in the SAD group showed significant differences in activities of the FPPFC (right: t=–2.93, P<.001; left: t=–2.25, P=.03) and the orbitofrontal cortex (OFC) (right: t=–2.10, P=.045; left: t=–2.21, P=.04) while watching video 2.

Conclusions: Activities of the FPPFC and OFC were associated with symptom reduction after VR treatment for SAD. Our study findings might provide a clue to understanding the mechanisms underlying VR treatment for SAD.

Trial Registration: Clinical Research Information Service (CRIS) KCT0003854; https://tinyurl.com/559jp2kp

(J Med Internet Res 2021;23(12):e31844) doi:10.2196/31844

KEYWORDS

anxiety; social anxiety disorder; virtual reality; fNIRS; brain activity; prefrontal cortex; effectiveness

Introduction

Social anxiety disorder (SAD) is a common psychiatric disease, with 8.4%-15% of the population worldwide diagnosed with it [1]. According to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5), SAD is characterized by fear or anxiety in situations in which people receive negative attention from others [2]. Individuals with SAD avoid social situations and have difficulties in maintaining interpersonal relationships, with serious impairment in academic, occupational, and social functions [3]. Comorbidities, such as depressive disorders, anxiety disorders, substance use disorder, obsessive compulsive disorder, and avoidant personality disorder, typically accompany SAD. Moreover, patients are likely to be single and unemployed and develop suicidal ideas, low self-esteem and a low quality of life, and chronic illnesses without treatment [4,5].

Effective pharmacological agents for SAD treatment include selective serotonin reuptake inhibitors, serotonin and norepinephrine reuptake inhibitors, benzodiazepines, and beta-adrenergic antagonists. Furthermore, nonpharmacological therapies, such as cognitive behavior therapy (CBT) or social skill training, and a combination of pharmacological and nonpharmacological therapies are considered treatment modalities [6-8]. Although these treatments are effective, patients with SAD may be reluctant to receive psychiatric medication, have difficulty in visiting the treatment room, and drop out of treatment to avoid social situations or spatial constraints. Thus, patients may not actively participate in treatments [9,10].

To overcome the limitations of conventional therapeutics, digital intervention strategies, such as virtual reality (VR), that incorporate advanced technology into conventional therapies have been developed. Since the 1990s, VR has been used in various clinical treatments in medicine. However, initially, it could not reproduce reality well due to technical limitations, but with gradual developments in digital technologies, such as graphics and sounds, the treatment effects of realistic VR have improved considerably [11-13].

Since Rothbaum et al [14] reported the effects of progressive exposure therapy using VR in acrophobia, many studies have reported the effects of VR in various psychiatric disorders, such as specific phobia, posttraumatic stress disorder, SAD, attention deficit hyperactivity disorder, autism spectrum disorder, schizophrenia, depressive disorders, anxiety disorders, eating disorders, addiction, and mild cognitive impairment [13,15-19]. VR therapy has numerous advantages because it enables self-directed treatment at home or in locations where the patient feels comfortable. Furthermore, VR motivates patients toward treatment, mitigating the limitations of conventional therapist-led therapy [15,20,21]. VR can depict various social situations that cannot be reproduced in real treatment spaces because of spatial or human constraints and allows modification of patient treatment for specific situations [22].

Although VR treatment for SAD has attracted considerable attention, it is still a novel treatment modality. Pertaub et al [23] identified that social anxiety can be induced by virtual situations, which yielded results similar to conventional exposure therapy. Robilard et al [24] reported improved detection of anxiety symptoms with VR-CBT in patients with SAD. Randomized controlled trials have reported the effects of VR exposure therapy in patients with SAD [25-27]. However, limited clinical studies have been performed for investigating the effectiveness of VR treatment, and most studies have only identified effectiveness based on psychological assessments that relied on patients' subjective expressions [28].

In this study, we designed an advanced therapeutic tool compared with existing VR treatments. During the sessions, the participants' own voices were recorded, and scenes were constructed with the participants' self-introduction in first person and third person to introduce participatory and interactive features that could be used therapeutically. The use of this VR program for measuring the psychological scale changes in patients with SAD has been reported earlier in a study [29]. In this study, we examined changes and responses in brain frontal region activities using functional near-infrared spectroscopy (fNIRS). fNIRS can be used to examine the functional activity in certain areas of the brain by measuring changes in the concentrations of oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HbR) in brain tissue to assess neurological activation. fNIRS is safe, portable, and easy to use compared with neuroimaging tools such as functional magnetic resonance imaging (fMRI) [30]. Previous neuroimaging studies have reported changes in the medial prefrontal cortex (MPFC), ventrolateral prefrontal cortex (VLPFC), amygdala, anterior cingulate cortex (ACC), and posterior cingulate cortex (PCC) in SAD. The changes in the prefrontal cortex (PFC) were reported before and after VR therapy and during treatment [31].

The purpose of this study is to identify the neural correlates of symptom improvement after VR treatment in SAD based on the efficacy of the VR program reported previously [29]. We used VR-derived video clips with first- or third-person interviews to observe changes in brain activity. The neural correlates of symptom improvement after VR treatment in patients with SAD were studied by analyzing and assessing changes in neuronal activities using fNIRS compared with healthy controls (HCs).

Methods

Study Participants

In total, 40 patients with SAD and 34 HCs matched for age, sex, and handedness were enrolled in the study. After data quality verification with fNIRS, we included 28 patients (19 [68%] women) diagnosed with primary SAD and 27 HCs (13 [48%] women) in the study. The SAD and HC groups were recruited from universities through online advertisements on bulletin boards. The inclusion criteria for the SAD group were as follows:

- Korean-speaking men or women aged between 19 and 31 years
- Satisfy DSM-IV criteria for SAD according to the Mini-International Neuropsychiatric Interview (MINI; a psychiatrist conducted MINI to select the SAD and HC groups in this study) [32]
- Psychotropic medication–naive patients without psychiatric comorbidities (excluding depressive and panic disorders)
- Patients not undergoing psychotherapy
- Patients without a history of neurological disorders
- Patients without a history of psychotic symptoms upon VR exposure
- Patients who experience epilepsy and are not vulnerable to visual stimuli

The exclusion criteria were as follows:

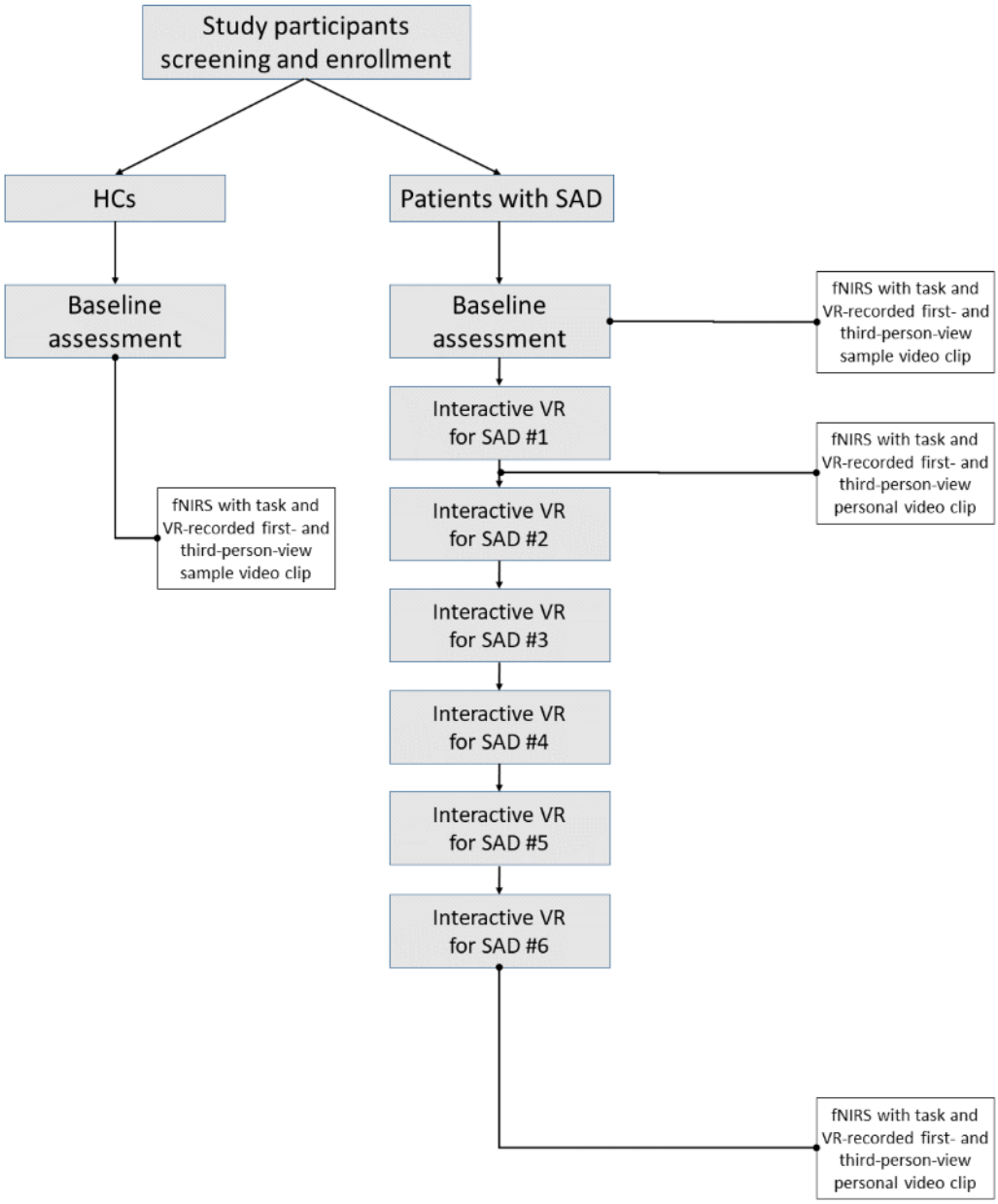
- History of intellectual disability or organic brain damage
- Individuals who experience psychotic symptoms upon VR experiences
- People vulnerable to visual stimuli
- People who are unsuitable for participation in fNIRS research (eg, people who cannot sustain the fNIRS test because of discomfort or anxiety in wearing fNIRS equipment or experiences of side effects, such as headache)

The HC group did not have preexisting neurological or psychiatric conditions. The Korean Social Avoidance and Distress (K-SAD) scale was prepared using an online questionnaire, and patients with a score of 82 or above were enrolled in the SAD group. All participants provided informed written consent after explanation of the study procedures and were required to pass an fNIRS scanning safety eligibility test. The study was approved by the Korea University Anam Hospital institutional review board (IRB no. 2018AN0377). The study was conducted in accordance with the Declaration of Helsinki ethical principles.

VR Therapy for SAD

An overview of the participatory VR treatment for SAD is available in previous reports [29,33]. Here, a similar participatory VR program was used for the treatment of patients with SAD. The VR program had an introduction, a core, and a finishing phase, with 3 levels depending on the difficulty of the core phase. In the introductory phase, participants learned how to use VR for relaxation and meditation. During the core phase, virtual situations were set up. In these virtual setups, several participants met to perform team tasks. Participants were exposed to social situations and introduced themselves to an audience. The level of difficulty was gradually increased, and participants were required to engage in discussions, and the activities were repeated. Each video clip was recorded in first- and third-person views [34]. The mirroring techniques in treatment were modified, and a hybrid treatment was used in the study. Namely, we attempted to maximize the effectiveness of treatment by consulting or discussing with participants by reconfirming VR treatment performed by themselves from the first- and third-person views. Each participant underwent 6 sessions of the treatment program. Participants were allowed to select the difficulty level of the program according to their comfort level after the first session. The researchers supervised each session to ensure the safety of the participants. The psychological scales were evaluated before the commencement of the VR program and after its completion. When the participants were watching the video clip in first and third person, we measured neuronal activities in the desired brain regions using fNIRS before VR treatment initialization and after the first and sixth sessions (Figure 1).

Figure 1. Overall schematic structure of the study. Patients with SAD and HCs were enrolled in a participatory and interactive VR treatment study. The HC group implemented fNIRS once by watching first- and third-person-view video clips at baseline. The SAD group implemented fNIRS 3 times in total at baseline, after the first and sixth sessions of VR treatment. fNIRS: functional near-infrared spectroscopy; HC: healthy control; SAD: social anxiety disorder; VR: virtual reality.



Clinical Symptom Assessments

The clinical symptoms of the participants were assessed using psychological scales. Subjects with SAD were scanned in 2 testing sessions, before and after VR therapy. We reported an analysis of psychological scale changes to evaluate the effectiveness of VR therapy for patients with SAD [29]. In this study, we studied the correlation between fNIRS results and psychological scales. Diagnostic assessment was measured using MINI [31]. Additionally, for social anxiety, we considered the Korean version of the Beck Anxiety Inventory (BAI) [35], and State-Trait Anxiety Inventory-X (STAI-X) [36]. Moreover, the Internalized Shame Scale (ISS) [37], the Post-Event Rumination Scale (PERS) [38], the Korean version of the Social Phobia Scale (K-SPS) [39], and the Korean version of the Social Interaction Anxiety Scale (K-SIAS) [39] were evaluated. Next, we also incorporated the Brief-Fear of Negative Evaluation (BFNE) scale [40], the K-SAD scale [41], and the Liebowitz Social Anxiety Scale (LSAS) [42].

For the HC group, fNIRS was evaluated once, whereas for the SAD group, it was measured 3 times (at baseline, after the first VR session, and after the sixth VR session). In each measurement, a 2-back task was conducted between the first- and third-person versions. We used a procedure similar to Wagner's task [43]. Working memory in the SAD group was studied previously using the N-back task, in which the SAD group exhibited lower accuracy and longer response times than the HC group [44]. Similarly, cognitive function was assessed using the N-back task, which measured working memory. To provide human-recorded first- and third-person views, fNIRS

was performed before and after the N-back operation to predict the expected therapeutic effect of VR-based psychotherapy. The 2-back task required the participants to determine whether the current stimulus was similar to the stimulus presented earlier in the trials. The task paradigm was composed of a block design that was repeated 3 times with alternate rest and task periods of 30 seconds.

fNIRS Data Analysis

The fNIRS system (NIRSIT, OBELAB Inc., Seoul, Republic of Korea), which monitors activation in the PFC, is a portable, wireless, wearable, and multichannel (48 channels) brain imaging system that covers the entire forehead. Wavelengths of 780 and 850 nm were used, and 48 regions of light source and detectors separated by 3 cm were considered. The weight of the device is 550 g, and the sampling rate is 8.138 Hz. To remove high-frequency noise due to environmental artifacts and low-frequency noise due to the blood circulation in the body, the fNIRS signal was filtered using a bandpass filter with a cutoff frequency of 0.005-0.1 Hz. Considering the signal-to-noise ratio, the signal quality of each channel was evaluated after filtering, and values less than 30 dB were eliminated as unreliable.

Using the wavelength-dependent hemoglobin absorption coefficient and differential path length coefficient, the modified Beer-Lambert law was applied to the filtered signal, and changes in oxyhemoglobin (ΔHbO_2) [45] and deoxyhemoglobin (ΔHbR) concentrations were extracted [46]. To account for the time-dependent change in hemodynamics, baseline correction was performed by subtracting the average from the calculated change in the oxygen concentration. Next, based on the Montreal Institute of Neurology standardization space (ICBM152), all 48 channels were grouped into 8 regions. After grouping each region of the PFC, the rejected channels identified during the channel rejection process were filled with the average ΔHbO_2 of the channels allowed within each region.

Statistical Analysis

The mean and standard deviation of demographic data, such as age and educational year, were calculated for both groups. After evaluating the variance equality through the Levene test, the difference in ΔHbO_2 between the HC and SAD groups was analyzed by conducting an independent t test. One-way repeated ANOVA and post hoc analysis were conducted to understand subject variation in the SAD group from baseline. After the first and sixth VR sessions, the Box test for equivalence check and the Mauchly test of sphericity were conducted to analyze fNIRS data. Data processing was performed using the NIRSIT package in MATLAB (version 2019b; MathWorks Inc., Natick, MA, USA). Finally, we used the Pearson correlation coefficient to identify the relationship between psychological scale scores and area-specific changes in brain activation. Additional statistical analyses were performed using IBM SPSS Statistics 21 (SPSS Inc., Chicago, IL, USA), where the criterion for statistical significance was set at $P<.05$.

Results

Demographic Data

In this study, we analyzed the data from 28 patients with SAD and 27 people in the HC group. The mean age (SD) for patients with SAD were 23.74 (3.55) years, and for the HC group, they were 23.18 (3.27) years. The mean (SD) educational years for the SAD and HC groups were 15.00 (1.88) and 14.63 (1.37), respectively. No difference was observed between the groups in terms of age and years of education.

The mean (SD) K-SAD scores were 106.68 (15.33) for the SAD group and 48.93 (17.10) for the HC group. Significant differences were observed in the severity of SAD symptoms (Table 1).

Table 1. Demographic and clinical data between SAD[a] patients and HCs[b].

Demographic and clinical data	Patients with SAD	HCs
Gender, n (%) male/n (%) female	28 (9 [32%]/19 [68%])	27 (14 [52%]/13 [48%])
Age in years, mean (SD)	23.74 (3.55)	23.18 (3.27)
Education years, mean (SD)	15.00 (1.88)	14.63 (1.37)
K-SAD[c] score[d], mean (SD)	106.68 (15.33)	48.93 (17.10)
STAI-S[e] score[d], mean (SD)	49.21 (10.17)	35.48 (7.81)
STAI-T[f] score[d], mean (SD)	53.21 (10.30)	34.19 (7.19)
ISS[g] score[d], mean (SD)	51.86 (16.70)	15.22 (9.74)
BFNE[h] score[d], mean (SD)	44.39 (8.91)	27.74 (7.35)
PERS[i] score[d], mean (SD)	46.96 (8.34)	32.81 (11.61)
LSASanx[j] score[d], mean (SD)	38.86 (12.66)	13.96 (10.56)

[a]SAD: social anxiety disorder.

[b]HC: healthy control.

[c]K-SAD: Korean Social Avoidance and Distress.

[d]$P<.01$.

[e]STAI-S: State-Trait Anxiety Inventory-State.

[f]STAI-T: State-Trait Anxiety Inventory-Trait.

[g]ISS: Internalized Shame Scale.

[h]BFNE: Brief-Fear of Negative Evaluation.

[i]PERS: Post-Event Rumination Scale.

[j]LSASanx: Liebowitz Social Anxiety Scale-anxiety.

Psychological Scale Scores of SAD and HC Groups

In this study, the scores of the SAD group were considerably higher than those of the HC group for all the tested scales that assessed associated anxiety symptoms (Table 1). In the SAD group, after VR treatment, the scores on the psychological scales for the SAD group decreased in all cases except 2, namely STAI-S (subscale of STAI, $P=.11$) and the Liebowitz Social Anxiety Scale-anxiety (LSASanx, subscale of LSAS, $P=.16$). (In a previous paper, we detailed the psychological scales used in the study and their purposes of assessment. Furthermore, we previously reported the changes in psychological scales according to VR treatment [29]).

Baseline PFC Activity in SAD and HC Group Measure Using fNIRS

We conducted fNIRS measurements for study participants before VR treatment to analyze brain activity in each region of the PFC, such as the dorsolateral prefrontal cortex (DLPFC), the VLPFC, the frontopolar prefrontal cortex (FPPFC), and the orbitofrontal cortex (OFC); see Figure 2. Brain activities among the SAD and HC groups were compared. A reduction in activation was measured by viewing first-person video clips (video 1), especially in the right FPPFC ($P<.05$). However, a negligible difference was observed between the two groups when the 2-back task was executed, and video clips in the third person were used for the treatment program (video 2). Next, while executing the 2-back task, activations in the left-brain areas of the HC group were lower than those of the SAD group, but the difference was not significant (Table 2).

Figure 2. Activity of the PFC in the HC and SAD groups measured by fNIRS at baseline. When fNIRS was performed while both groups watched video 1, the activity of the right FPPFC was more decreased in the SAD group than in the HC group (P<.05). However, when implementing the 2-back task and watching video 2, there was no significant difference in the activity of the PFC between both groups. fNIRS: functional near-infrared spectroscopy; FPPFC: frontopolar prefrontal cortex; HC: healthy control; PFC: prefrontal cortex; SAD: social anxiety disorder.

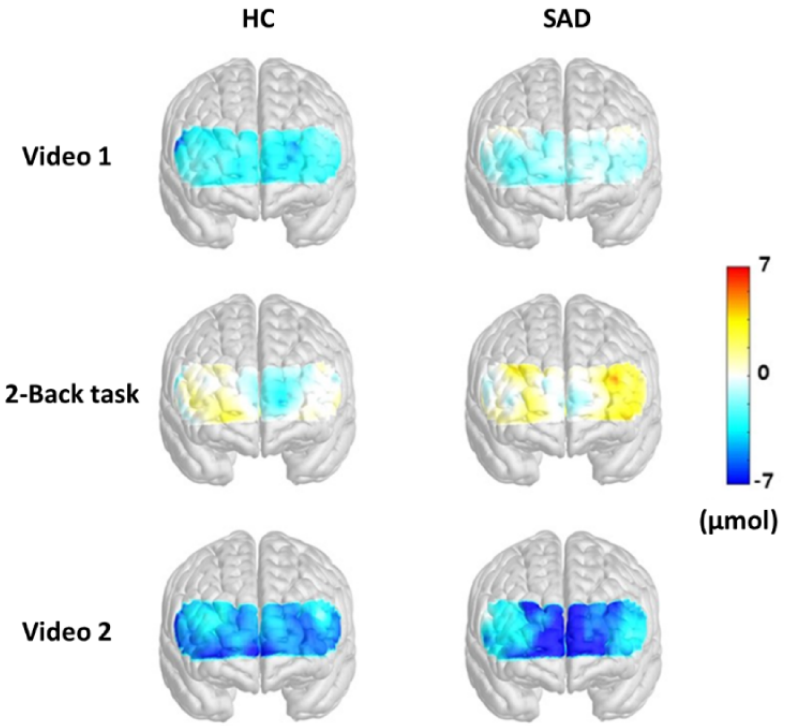


Table 2. Comparison of the activation of the PFC[a] measured by fNIRS[b] between SAD[c] patients and HCs[d] at baseline.

PFC region	*t*	*df*	*P* value
Video 1			
DLPFC[e] (R[f])	–1.8358	53	.07
VLPFC[g] (R)	–0.9965	53	.32
FPPFC[h] (R)	–2.0141	53	*.049* [i]
OFC[j] (R)	–0.4206	53	.68
DLPFC (L[k])	–1.5172	53	.14
VLPFC (L)	–0.9989	53	.32
FPPFC (L)	–1.5577	53	.13
OFC (L)	–1.6796	53	.10
2-back task			
DLPFC (R)	–0.1724	53	.86
VLPFC (R)	0.3430	53	.73
FPPFC (R)	–0.5709	53	.57
OFC (R)	0.0677	53	.95
DLPFC (L)	–1.9085	53	.06
VLPFC (L)	–1.7692	53	.08
FPPFC (L)	–1.8762	53	.07
OFC (L)	–1.8244	53	.07
Video 2			
DLPFC (R)	0.2410	53	.81
VLPFC (R)	0.1600	53	.87
FPPFC (R)	0.6791	53	.50
OFC (R)	–0.3050	53	.76
DLPFC (L)	–0.1003	53	.92
VLPFC (L)	–0.2931	53	.20
FPPFC (L)	0.5063	53	.61
OFC (L)	–1.8244	53	.84

[a]PFC: prefrontal cortex.

[b]fNIRS: functional near-infrared spectroscopy.

[c]SAD: social anxiety disorder.

[d]HC: healthy control.

[e]DLPFC: dorsolateral prefrontal cortex.

[f]R: right side.

[g]VLPFC: ventrolateral prefrontal cortex.

[h]FPPFC: frontopolar prefrontal cortex.

[i]*P* values in italics are significant ($P<.05$).

[j]OFC: orbitofrontal cortex.

[k]L: left side.

Changes in PFC Activity in SAD After VR Treatment Using fNIRS

We compared the brain activities measured by fNIRS (1) at baseline with those obtained after the first and sixth sessions of treatment in patients with SAD and (2) those obtained after the first session with those obtained after the sixth session. We observed significant differences in the right FPPFC ($P=.01$) and OFC ($P=.045$) and the left FPPFC ($P=.03$) and OFC ($P=.04$) when the treatment plan included video 2 for patients with SAD (Table 3 and Figure 3).

Table 3. Comparison of the activation of the PFC[a] measured through fNIRS[b] before and after VR[c] treatment in SAD[d] patients.

PFC region	*df*	Baseline vs session 1		Baseline vs session 6	
		t	*P* value	*t*	*P* value
Video 1					
DLPFC[e] (R[f])	27	0.2855	.78	–0.3951	.70
VLPFC[g] (R)	27	–0.0279	.98	–0.6768	.50
FPPFC[h] (R)	27	–0.4071	.69	–1.2572	.22
OFC[i] (R)	27	–1.9413	.06	–1.5350	.14
DLPFC (L[j])	27	–0.7767	.44	–0.7083	.48
VLPFC (L)	27	–1.8585	.07	–1.5250	.14
FPPFC (L)	27	–0.8758	.39	–0.9972	.33
OFC (L)	27	–0.1165	.91	–0.3891	.70
2-back task					
DLPFC (R)	27	1.1630	.25	0.5698	.57
VLPFC (R)	27	1.7643	.09	1.0058	.32
FPPFC (R)	27	0.9091	.37	1.1983	.24
OFC (R)	27	1.7939	.08	2.4581	*.02* [k]
DLPFC (L)	27	0.2466	.81	0.3295	.74
VLPFC (L)	27	0.9830	.33	0.4685	.64
FPPFC (L)	27	0.3272	.75	0.7516	.46
OFC (L)	27	1.7245	.10	2.0201	.05
Video 2					
DLPFC (R)	27	–1.7973	.08	–1.7163	.10
VLPFC (R)	27	–1.2856	.21	–1.3256	.20
FPPFC (R)	27	–3.1821	*.004* [l]	–2.9260	*.007* [l]
OFC (R)	27	–2.8014	*.009* [l]	–2.0984	*.045* [k]
DLPFC (L)	27	–1.8807	.07	–1.8464	.08
VLPFC (L)	27	–0.9249	.36	–1.4879	.15
FPPFC (L)	27	–2.3205	*.028* [k]	–2.2514	*.03* [k]
OFC (L)	27	–2.6050	*.015* [k]	–2.2136	*.035* [k]

[a]PFC: prefrontal cortex.

[b]fNIRS: functional near-infrared spectroscopy.

[c]VR: virtual reality.

[d]SAD: social anxiety disorder.

[e]DLPFC: dorsolateral prefrontal cortex.

[f]R: right side.

[g]VLPFC: ventrolateral prefrontal cortex.

[h]FPPFC: frontopolar prefrontal cortex.

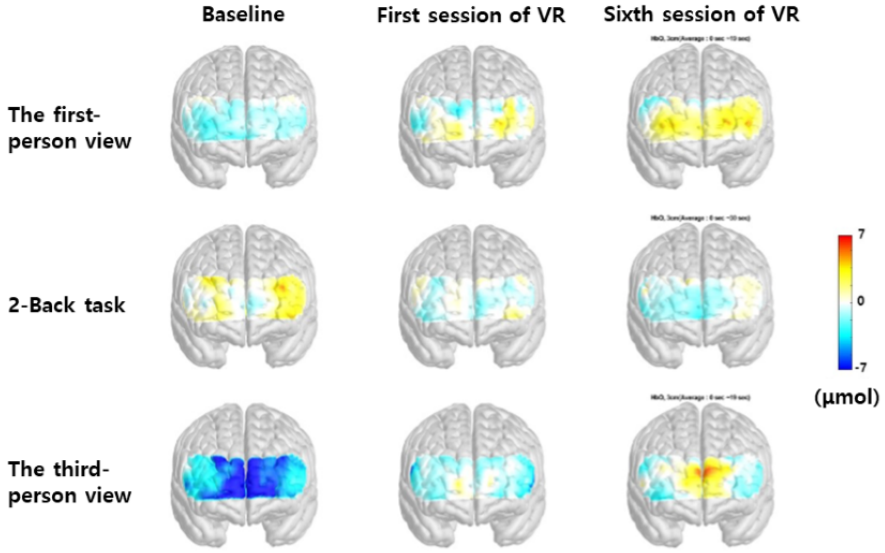
[i]OFC: orbitofrontal cortex.

[j]L: left side.

[k]*P* values in italics are significant ($P<.05$).

[l]*P* values in italics are significant ($P<.01$).

Figure 3. Activity of the PFC of patients with SAD at baseline, after the first VR session and after the sixth VR session. Comparing between baseline and after the first VR session, when watching video 2, there were significant differences in activity in the right FPPFC (P<0.01) and OFC (P<0.01) and in the left FPPFC (P<0.05) and OFC (P<0.05). Comparing between baseline and after the sixth VR session, when watching video 2, there were significant differences in activity in the right FPPFC (P<0.01) and OFC (P<0.05) and in the left FPPFC (P<0.05) and OFC (P<0.05). When patients with SAD were performing the 2-back task, there was a significant difference in activity in the right OFC (P<0.05) between baseline and after the sixth VR session. There was no significant difference in PFC activity when watching video 1. FPPFC: frontopolar prefrontal cortex; OFC: orbitofrontal cortex; PFC: prefrontal cortex; SAD: social anxiety disorder; VR: virtual reality.



Correlations Between Brain Activities and Psychological Scale Results

We analyzed the correlations between psychological scales and brain activities through fNIRS before and after VR treatment in patients with SAD. The STAI-S scores showed significant correlations with the activities of the right and left FPPFC, the right and left OFC, and the left DLPFC. Furthermore, the STAI-T scores were significantly correlated with the right and left OFC, and the ISS was significantly correlated with the right and left FPPFC and the right and left OFC in video 1. We also determined that STAI-S scores were significantly correlated with the right and left VLPFC, the right FPPFC, and the left DLPFC; the ISS was significantly correlated with the right and left DLPFC, the left VLPFC, the right and left FPPFC, and the right and left OFC; and the BFNE was significantly correlated with the right VLPFC in video 2. IN addition, PERS was significantly correlated with the right OFC, the K-SAD scale was significantly correlated with the left OFC, and LSASanx was significantly correlated with the right DLPFC, VLPFC, FPPFC and the right and left OFC while executing the 2-back task (Table 4).

Table 4. Correlations between brain activities and results of psychological scales.

Psychological scale	DLPFC[a] (R[b])	VLPFC[c] (R)	FPPFC[d] (R)	OFC[e] (R)	DLPFC (L[f])	VLPFC (L)	FPPFC (L)	OFC (L)
Video 1								
BAI[g]	–.162	–.103	–.161	–.149	–.191	–.153	–.120	–.082
STAI-S[h]	–.321	–.199	*–.550*[i]	*–.527*[i]	*–.436*[j]	–.304	*–.466*[j]	*–.483*[i]
STAI-T[k]	–.109	–.097	–.329	*-.414* [j]	–.259	–.241	–.296	*–.393*[j]
ISS[l]	–.297	–.175	*–.388*[j]	*–.400*[j]	–.307	–.191	*–.399*[j]	*–.439*[j]
PERS[m]	.075	.086	.213	.075	–.032	–.038	.031	–.070
SPS[n]	–.060	.053	–.117	–.117	–.088	–.009	–.113	–.060
SIAS[o]	.152	.127	.092	.090	.144	.083	.131	.132
BFNE[p]	.221	.208	.188	.350	.272	.135	.226	.269
K-SAD[q]	.086	.098	–.023	.246	.132	.187	.116	.227
LSASanx[r]	–.172	–.221	–.179	–.028	–.159	–.133	–.191	–.141
LSASavo[s]	–.073	–.097	–.021	.022	–.035	–.008	–.015	–.059
2-back task								
BAI	.172	–.035	.067	.003	.099	.014	.097	.122
STAI-S	.192	–.082	.144	.089	.044	–.182	.168	.065
STAI-T	.170	–.039	.153	.090	.078	–.022	.112	.052
ISS	.032	.164	.258	.037	–.014	–.096	.312	–.027
PERS	–.355	–.322	–.274	*–.426*[j]	–.091	–.106	–.318	–.187
SPS	–.155	–.360	–.028	–.269	.082	–.136	.026	–.280
SIAS	–.095	–.312	–.237	–.375	–.014	–.107	–.252	–.340
BFNE	–.102	.054	–.003	–.184	–.171	–.161	.103	–.255
K-SAD	–.027	–.182	–.178	–.230	–.009	–.133	–.115	*–.377*[j]
LSASanx	*–.457*[j]	*–.375*[j]	*–.395*[j]	*–.443*[j]	–.230	–.232	–.338	*–.474*[j]
LSASavo	–.312	–.240	–.303	–.335	–.252	–.192	–.284	–.369
Video 2								
BAI	.115	.096	.099	.146	.051	.071	.074	.118
STAI-S	–.359	*–.402*[j]	–.382[j]	–.234	*–.482*[i]	*–.460*[j]	–.325	–.357
STAI-T	–.091	–.157	–.147	–.035	–.257	–.230	–.136	–.181
ISS	*–.413*[j]	–.357	*–.426*[j]	*–.403*[j]	*–.452*[j]	*–.381*[j]	*–.432*[j]	*–.381*[j]
PERS	–.013	–.040	.081	–.002	.036	.062	–.052	.039
SPS	.041	.027	–.002	.020	.048	.007	.089	.168
SIAS	–.015	–.091	–.114	–.062	.068	.103	.060	.174
BFNE	.363	*.484* [i]	.183	.175	.347	.370	.208	.234
K-SAD	.057	.013	–.148	–.072	.039	.072	–.028	.011
LSASanx	–.062	–.127	–.162	–.038	–.149	–.096	–.109	–.040
LSASavo	.045	–.007	–.059	–.009	–.018	.044	–.047	.024

[a]DLPFC: dorsolateral prefrontal cortex.

[b]R: right side.

[c]VLPFC: ventrolateral prefrontal cortex.
[d]FPPFC: frontopolar prefrontal cortex.
[e]OFC: orbitofrontal cortex.
[f]L: left side.
[g]BAI: Beck Anxiety Inventory.
[h]STAI-S: State-Trait Anxiety Inventory-State.
[i]*P* values in italics are significant (*P*<.01).
[j]*P* values in italics are significant (*P*<.05).
[k]STAI-T: State-Trait Anxiety Inventory-Trait.
[l]ISS: Internalized Shame Scale.
[m]PERS: Post-Event Rumination Scale.
[n]SPS: Social Phobia Scale.
[o]SIAS: Social Interaction Anxiety Scale.
[p]BFNE: Brief-Fear of Negative Evaluation.
[q]K-SAD: Korean Social Avoidance and Distress.
[r]LSASanx: Liebowitz Social Anxiety Scale-anxiety.
[s]LSASavo: Liebowitz Social Anxiety Scale-avoidance.

Discussion

Principal Findings

To the best of our knowledge, this study is the first to determine the effects of VR therapy on SAD by using fNIRS measurements. From this study, we found that changes in activity in specific brain regions, such as the FPPFC and OFC, are closely related to VR treatment. In addition, we developed a VR program to enhance the efficacy of the treatment by creating first- and third-person views of exposure situations. In the first-person view, the patients directly present content to an audience, and in the third-person view, the patients visualize themselves speaking to an audience objectively.

The effects of VR treatment on patients with SAD were studied by measuring the changes in PFC brain activity using fNIRS. We were able to find that the activation of the right FPPFC was higher in the SAD group than in the HC group when watching video 1. This can be understood in a similar context as the blood flow of the MPFC decreases in situations in which anxiety is predicted in healthy individuals [47,48]. A positron emission tomography (PET) study reported that the blood flow in the right DLPFC, left inferior temporal cortex, and left amygdaloid-hippocampal region increases in patients with SAD [48]. However, using fNIRS analysis, we did not observe any difference in activation between the right DLPFC among the study groups. Previously, an fNIRS scan revealed that the activation change in the VLPFC in patients with SAD was less than that in the HC group during a verbal fluency task [49]. Here, the activation change in most regions of the left PFC was low and not significant.

In this study, we observed changes in the prefrontal area before and after VR treatment. Thus, activation in the right FPPFC and OFC and the left FPPFC and OFC significantly increased when patients with SAD were watching the video in the third-person view. By using PET after pharmacological treatments, Evans et al [50] reported that the regional brain metabolism of the ventromedial prefrontal cortex (VMPFC) in patients with SAD increases. Further, Hiser et al [51] reported that the activation of the anterior VMPFC increases upon pharmacological treatment or psychotherapy. We obtained similar results to these previous studies. However, the metabolism in the medial dorsal PFC decreased in patients with SAD after pharmacological treatment [52].

We obtained positive self-referential stimuli (positive words), neutral self-referential stimuli, and negative self-referential stimuli (negative words) after VR treatment to identify the activation changes in the brain region measured through fMRI in a previous study [33]. Therefore, patients with SAD exhibited increased activation in the right PCC/precuneus, lingual gyrus, left inferior temporal gyrus, precentral gyrus, and postcentral gyrus for positive self-referential stimuli. Patients with SAD also revealed increased activation in the left middle occipital gyrus, parahippocampus, left Rolandic operculum, and left caudate nucleus for negative self-referential stimuli. In this study, we observed the brain activity of participants watching the video clips, unlike the previous study, and determined the neural correlates by fNIRS, which were used to measure the activation of the prefrontal area, unlike fMRI, which can be used to measure the activation of the whole brain area.

Lastly, we identified correlations between psychological states in patients with SAD with brain areas activated upon VR therapy. The STAI-T scores and OFC activation, and ISS scores and prefrontal region activation (except the right VLPFC) were correlated. Studies have reported the correlation between the severity of SAD symptoms and the activation of certain brain areas. Klumpp et al [53] reported that improvements in the symptoms are correlated with the increase in the activation of brain areas, such as the medial orbitofrontal and dorsomedial frontal gyrus. Marin et al [54] reported that hypoactivation of the VMPFC is correlated with the severity of anxiety. Based on fNIRS analysis, Yokoyama et al [49] reported that changes in VLPFC activation are negatively correlated with social fear and activation of the left DLPFC is positively correlated with social anxiety [55].

Studies have reported that amygdala activation is high in patients with SAD [48,56]. Here, we could not study amygdala activation, because measuring subcortical brain regions with

fNIRS is difficult. The VMPFC and the VLPFC are known to be associated with the amygdala in patients with SAD. Generally, the VMPFC regulates social functions and inhibits the amygdala to regulate the fear response [51]. Thus, increased activation of the ventromedial prefrontal regions from social anxiety treatments is beneficial. In addition, the degree of functional connectivity between the VLPFC and the amygdala is negatively correlated with the severity of anxiety [57]. This result supports our results that activation of the VLPFC and the degree of anxiety treatment are correlated.

The blood flow in the right DLPFC decreased in patients with SAD during the repetition task [48]. We observed that activation of the right OFC during the 2-back task in our study increased after treatment, which contradicts the results of some previous studies. Further, regional cerebral blood flow of the rhinal cortex, amygdala, hippocampal region, and parahippocampal region decreases in patients with SAD during task performance. The regional cerebral blood flow of the middle frontal cortex and dorsal ACC changes in patients with SAD, particularly in patients who effectively perform the task [58]. We obtained similar results in our study.

Differences were observed in brain activation in high anxiety–provoking situations rather than in low anxiety–provoking situations [59]. Patients with SAD were more anxious during the first-person view than during the third-person view, which resulted in differences in activation in the HC group. In addition, the changes in brain activation appeared earlier when the video in the third-person view was used, because the patients felt less anxious. However, the effects of the treatment appeared later when the video in the first-person view was used, probably because the patients felt more anxious. Therefore, the effectiveness of the treatment modality should be verified by conducting more treatment sessions.

Limitations

This study had some limitations. First, the sample size was relatively small. However, our analysis was thorough, and we were able to investigate functional brain activity after VR therapy in patients with SAD. Second, the number of sessions for VR treatment was small, and the follow-up period was relatively short. We observed the changes in psychological scale scores and brain activities measured through fNIRS after VR treatment. However, after the sixth session of treatment, the anxiety levels observed on most psychological scales in the SAD group were higher than in the HC group. This could be because changes in brain activity due to VR treatment were not prominent as yet. Perhaps more treatment sessions are required. Thus, more VR treatment sessions should be conducted for a larger sample size to observe prominent therapeutic benefits. Third, we could not measure the brain activity in real time. Applying electroencephalography to a VR head-mounted display (HMD) has enabled measurement of brain waves [60]. However, we were not able to apply fNIRS and VR HMD because of several technical limitations. Lastly, we could not observe changes in brain areas other than the PFC (especially amygdala), owing to mechanical limitations of portable fNIRS. Although fMRI has the advantage of being able to perform comprehensive and in-depth measurement of the entire brain function, it is difficult to apply various tasks and has a high economic burden, so there are obvious limitations in actual clinical application. However, in terms of cost and usability, portal fNIRS has a high possibility of being applied to actual clinical practice in combination with VR, so it is worthy of overcoming the disadvantages of measurement compared to fMRI.

Conclusion

We conducted VR therapy on patients with SAD to prove the treatment effects after developing a VR treatment program that was exposed to social situations in the first- and third-person views. We identified the brain activity changes in the FPPFC and OFC in patients with SAD after VR treatment in this study, which can provide functional clues in specific brain area for VR treatment. In addition, we found that the STAI-T scores were correlated with the activity of the OFC and the ISS scores were correlated with most of the PFC except the right VLPFC. This can be a clue that the change after treatment in some psychological domains in SAD is related to a specific PFC region. Future studies are required to identify and confirm the neural correlates of VR treatment in patients with SAD for sufficient periods in large sample sizes.

Acknowledgments

This work was supported by the National Research Foundation (NRF) of Korea grant funded by the Ministry of Science and Information and Communications Technology (MSIT), Government of Korea (NRF-2020R1C1C1007463 and NRF-2021R1A5A8032895) and the Information and Communications Technology (ICT) & Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2017M3C1B6070978).

Authors' Contributions

HL, JKC, DJ, JWH, and CHC wrote the first draft of the manuscript. JKC, DJ, JWH, and CHC participated in data collection. All the authors edited the manuscript versions. All the authors were also involved in the interpretation of the results and read, commented on, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Heimberg RG, Hofmann SG, Liebowitz MR, Schneier FR, Smits JAJ, Stein MB, et al. Social anxiety disorder in DSM-5. Depress Anxiety 2014 Jun;31(6):472-479. [doi: 10.1002/da.22231] [Medline: 24395386]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013.
3. Bögels SM, Alden L, Beidel DC, Clark LA, Pine DS, Stein MB, et al. Social anxiety disorder: questions and answers for the DSM-V. Depress Anxiety 2010 Feb;27(2):168-189. [doi: 10.1002/da.20670] [Medline: 20143427]
4. Cuijpers P, van Straten A. Improving outcomes in social anxiety disorder. Lancet Psychiatry 2014 Oct;1(5):324-326. [doi: 10.1016/s2215-0366(14)70351-7]
5. Keller MB. The lifelong course of social anxiety disorder: a clinical perspective. Acta Psychiatr Scand Suppl 2003(417):85-94. [doi: 10.1034/j.1600-0447.108.s417.6.x] [Medline: 12950439]
6. Rodebaugh TL, Holaway RM, Heimberg RG. The treatment of social anxiety disorder. Clin Psychol Rev 2004 Nov;24(7):883-908. [doi: 10.1016/j.cpr.2004.07.007] [Medline: 15501560]
7. Bandelow B. Current and novel psychopharmacological drugs for anxiety disorders. Adv Exp Med Biol 2020;1191:347-365. [doi: 10.1007/978-981-32-9705-0_19] [Medline: 32002937]
8. Apolinário-Hagen J, Drüge M, Fritsche L. Cognitive behavioral therapy, mindfulness-based cognitive therapy and acceptance commitment therapy for anxiety disorders: integrating traditional with digital treatment approaches. In: Kim YK, editor. Anxiety Disorders: Rethinking and Understanding Recent Discoveries. Singapore: Springer; 2020:291-329.
9. Rose GM, Tadi P. Social Anxiety Disorder. Treasure Island, FL: StatPearls; 2021.
10. Olfson M, Guardino M, Struening E, Schneier FR, Hellman F, Klein DF. Barriers to the treatment of social anxiety. Am J Psychiatry 2000 Apr;157(4):521-527. [Medline: 10739410]
11. Rizzo AS, Koenig ST, Talbot TB. Clinical virtual reality: emerging opportunities for psychiatry. Focus (Am Psychiatr Publ) 2018 Jul;16(3):266-278 [FREE Full text] [doi: 10.1176/appi.focus.20180011] [Medline: 31975921]
12. Lee JM, Ku JH, Jang DP, Kim DH, Choi YH, Kim IY, et al. Virtual reality system for treatment of the fear of public speaking using image-based rendering and moving pictures. Cyberpsychol Behav 2002 Jun;5(3):191-195. [doi: 10.1089/109493102760147169] [Medline: 12123239]
13. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
14. Rothbaum BO, Hodges LF, Kooper R, Opdyke D, Williford JS, North M. Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. Am J Psychiatry 1995 Apr;152(4):626-628. [doi: 10.1176/ajp.152.4.626] [Medline: 7694917]
15. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: a systematic review of recent evidence. Psychiatry Res 2016 Feb 28;236:189-195. [doi: 10.1016/j.psychres.2016.01.015] [Medline: 26795129]
16. Oing T, Prescott J. Implementations of virtual reality for anxiety-related disorders: systematic review. JMIR Serious Games 2018 Nov 07;6(4):e10965 [FREE Full text] [doi: 10.2196/10965] [Medline: 30404770]
17. Zeng N, Pope Z, Lee JE, Gao Z. Virtual reality exercise for anxiety and depression: a preliminary review of current research in an emerging field. J Clin Med 2018 Mar 04;7(3):42 [FREE Full text] [doi: 10.3390/jcm7030042] [Medline: 29510528]
18. Vass E, Fekete Z, Lencse L, Ecseri M, Kis B, Szekrényes-Varga Á, et al. Treatment of Theory of Mind deficits in schizophrenia by using Virtual Reality, the VR-ToMIS training (Virtual Reality based Theory of Mind Intervention in Schizophrenia). Psychiatr Hung 2019;34(3):287-299. [Medline: 31570660]
19. Lalonde G, Henry M, Drouin-Germain A, Nolin P, Beauchamp MH. Assessment of executive function in adolescence: a comparison of traditional and virtual reality tools. J Neurosci Methods 2013 Sep 30;219(1):76-82. [doi: 10.1016/j.jneumeth.2013.07.005] [Medline: 23867080]
20. Didehbani N, Allen T, Kandalaft M, Krawczyk D, Chapman S. Virtual reality social cognition training for children with high functioning autism. Comput Hum Behav 2016 Sep;62:703-711. [doi: 10.1016/j.chb.2016.04.033]
21. Kellmeyer P. Neurophilosophical and ethical aspects of virtual reality therapy in neurology and psychiatry. Camb Q Healthc Ethics 2018 Oct;27(4):610-627. [doi: 10.1017/S0963180118000129] [Medline: 30198465]
22. Roy S, Klinger E, Légeron P, Lauer F, Chemin I, Nugues P. Definition of a VR-based protocol to treat social phobia. Cyberpsychol Behav 2003 Aug;6(4):411-420. [doi: 10.1089/109493103322278808] [Medline: 14511454]
23. Pertaub DP, Slater M, Barker C. An experiment on fear of public speaking in virtual reality. Stud Health Technol Inform 2001;81:372-378. [Medline: 11317771]
24. Robillard G, Bouchard S, Dumoulin S, Guitard T, Klinger E. Using virtual humans to alleviate social anxiety: preliminary report from a comparative outcome study. Stud Health Technol Inform 2010;154:57-60. [Medline: 20543270]
25. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, et al. Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. J Consult Clin Psychol 2013 Oct;81(5):751-760. [doi: 10.1037/a0033559] [Medline: 23796315]

26. Kampmann IL, Emmelkamp PMG, Hartanto D, Brinkman W, Zijlstra BJH, Morina N. Exposure to virtual social interactions in the treatment of social anxiety disorder: a randomized controlled trial. Behav Res Ther 2016 Feb;77:147-156. [doi: 10.1016/j.brat.2015.12.016] [Medline: 26752328]
27. Bouchard S, Dumoulin S, Robillard G, Guitard T, Klinger, Forget H, et al. Virtual reality compared with exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. Br J Psychiatry 2017 Apr;210(4):276-283. [doi: 10.1192/bjp.bp.116.184234] [Medline: 27979818]
28. Wechsler TF, Kümpers F, Mühlberger A. Inferiority or even superiority of virtual reality exposure therapy in phobias? A systematic review and quantitative meta-analysis on randomized controlled trials specifically comparing the efficacy of virtual reality exposure to gold standard exposure in agoraphobia, specific phobia, and social phobia. Front Psychol 2019;10:1758 [FREE Full text] [doi: 10.3389/fpsyg.2019.01758] [Medline: 31551840]
29. Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S, et al. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. Ann N Y Acad Sci 2020 Mar;1464(1):5-29 [FREE Full text] [doi: 10.1111/nyas.13948] [Medline: 30085354]
30. Yoon H, Seo EH, Kim J, Choo IH. Neural correlates of self-referential processing and their clinical implications in social anxiety disorder. Clin Psychopharmacol Neurosci 2019 Feb 28;17(1):12-24 [FREE Full text] [doi: 10.9758/cpn.2019.17.1.12] [Medline: 30690936]
31. Yoo SW, Kim YS, Noh JS, Oh KS, Kim CH, NamKoong K, et al. Validity of Korean version of the MINI-international neuropsychiatric interview. Anxiety Mood 2006;2(1):50-55.
32. Kim H, Lee S, Jung D, Hur J, Lee H, Lee S, et al. Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: longitudinal questionnaire study. J Med Internet Res 2020 Oct 06;22(10):e23024 [FREE Full text] [doi: 10.2196/23024] [Medline: 33021481]
33. Hur J, Shin H, Jung D, Lee H, Lee S, Kim GJ, et al. Virtual reality-based psychotherapy in social anxiety disorder: fMRI study using a self-referential task. JMIR Ment Health 2021 Apr 14;8(4):e25731 [FREE Full text] [doi: 10.2196/25731] [Medline: 33851931]
34. Cho CH. Social Anxiety Disorder VR treatment. 2019. URL: https://www.youtube.com/channel/UCCpzKK_qEqz8LRfxl_TfGXw?view_as=subscriber [accessed 2021-12-08]
35. Yook SP, Kim ZS. A clinical study on the Korean version of Beck Anxiety Inventory: comparative study of patient and non-patient. Korean J Clin Psychol 1997;16(1):185-197.
36. Kim JT, Shin DK. A study on the standardization of STAI for Korea. N Engl J Med 1978;21(11):69-75.
37. Lee IS, Choi HR. Assessment of shame and its relationship with maternal attachment, hypersensitive narcissism and loneliness. Korean J Couns Psychother 2005;17(3):651-670.
38. Im S, Kwon SM, Choi H. Korean J Clin Psychol 2007 Feb;26(1):39-56. [doi: 10.15842/kjcp.2007.26.1.003]
39. Kim HS. Memory Bias in Subtypes of Social Phobia (masters thesis). Seoul, South Korea: Seoul National University; 2001.
40. Hong Y, Moon J, Jo H. A validation study of the Korean version of the Fear of Negative Evaluation Scale-2. Korean J Psychol 2011;30:117-134.
41. Lee JY, Choi CH. A study of the reliability and the validity of the Korean versions of social phobia scales (K-SAD, K-FNE). Korean J Clin Psychol 1997;16(2):251-264.
42. Yu E, Ahn CY, Hwan PK. Factor structure and diagnostic efficiency of a Korean version of the Liebowitz Social Anxiety Scale. Korean J Clin Psychol 2007 Feb;26(1):251-270. [doi: 10.15842/kjcp.2007.26.1.015]
43. Wagner G, Koch K, Schachtzabel C, Peikert G, Schultz CC, Reichenbach JR, et al. Self-referential processing influences functional activation during cognitive control: an fMRI study. Soc Cogn Affect Neurosci 2013 Oct;8(7):828-837 [FREE Full text] [doi: 10.1093/scan/nss074] [Medline: 22798398]
44. Judah MR, Grant DM, Lechner WV, Mills AC. Working memory load moderates late attentional bias in social anxiety. Cogn Emot 2013;27(3):502-511. [doi: 10.1080/02699931.2012.719490] [Medline: 22963405]
45. Delpy DT, Cope M, van der Zee P, Arridge S, Wray S, Wyatt J. Estimation of optical pathlength through tissue from direct time of flight measurement. Phys Med Biol 1988 Dec;33(12):1433-1442. [doi: 10.1088/0031-9155/33/12/008] [Medline: 3237772]
46. Sun J, Liu F, Wang H, Yang A, Gao C, Li Z, et al. Connectivity properties in the prefrontal cortex during working memory: a near-infrared spectroscopy study. J Biomed Opt 2019 Mar;24(5):1-7 [FREE Full text] [doi: 10.1117/1.JBO.24.5.051410] [Medline: 30900431]
47. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-7 concl. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
48. Mathew SJ, Ho S. Etiology and neurobiology of social anxiety disorder. J Clin Psychiatry 2006;67 Suppl 12:9-13. [Medline: 17092190]
49. Yokoyama C, Kaiya H, Kumano H, Kinou M, Umekage T, Yasuda S, et al. Dysfunction of ventrolateral prefrontal cortex underlying social anxiety disorder: a multi-channel NIRS study. Neuroimage Clin 2015;8:455-461 [FREE Full text] [doi: 10.1016/j.nicl.2015.05.011] [Medline: 26106570]

50. Evans KC, Simon NM, Dougherty DD, Hoge EA, Worthington JJ, Chow C, et al. A PET study of tiagabine treatment implicates ventral medial prefrontal cortex in generalized social anxiety disorder. Neuropsychopharmacology 2009 Jan;34(2):390-398. [doi: 10.1038/npp.2008.69] [Medline: 18536708]
51. Hiser J, Koenigs M. The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. Biol Psychiatry 2018 Apr 15;83(8):638-647 [FREE Full text] [doi: 10.1016/j.biopsych.2017.10.030] [Medline: 29275839]
52. Doruyter A, Dupont P, Taljaard L, Stein DJ, Lochner C, Warwick JM. Resting regional brain metabolism in social anxiety disorder and the effect of moclobemide therapy. Metab Brain Dis 2018 Apr;33(2):569-581. [doi: 10.1007/s11011-017-0145-7] [Medline: 29101601]
53. Klumpp H, Fitzgerald DA, Phan KL. Neural predictors and mechanisms of cognitive behavioral therapy on threat processing in social anxiety disorder. Prog Neuropsychopharmacol Biol Psychiatry 2013 Aug 01;45:83-91 [FREE Full text] [doi: 10.1016/j.pnpbp.2013.05.004] [Medline: 23665375]
54. Marin M, Zsido RG, Song H, Lasko NB, Killgore WDS, Rauch SL, et al. Skin conductance responses and neural activations during fear conditioning and extinction recall across anxiety disorders. JAMA Psychiatry 2017 Jun 01;74(6):622-631 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0329] [Medline: 28403387]
55. Nishiyama Y, Okamoto Y, Kunisato Y, Okada G, Yoshimura S, Kanai Y, et al. fMRI study of social anxiety during social ostracism with and without emotional support. PLoS One 2015;10(5):e0127426 [FREE Full text] [doi: 10.1371/journal.pone.0127426] [Medline: 26000902]
56. Stein MB, Stein DJ. Social anxiety disorder. Lancet 2008 Mar;371(9618):1115-1125. [doi: 10.1016/s0140-6736(08)60488-2]
57. Fitzgerald JM, Klumpp H, Langenecker S, Phan KL. Transdiagnostic neural correlates of volitional emotion regulation in anxiety and depression. Depress Anxiety 2019 May;36(5):453-464 [FREE Full text] [doi: 10.1002/da.22859] [Medline: 30408261]
58. Doruyter AG, Dupont P, Stein DJ, Lochner C, Warwick JM. Nuclear neuroimaging in social anxiety disorder: a review. J Nucl Med 2018 Dec 05;59(12):1794-1800 [FREE Full text] [doi: 10.2967/jnumed.118.212795] [Medline: 30291197]
59. Glassman LH, Kuster AT, Shaw JA, Forman EM, Izzetoglu M, Matteucci A, et al. The relationship between dorsolateral prefrontal activation and speech performance-based social anxiety using functional near infrared spectroscopy. Brain Imaging Behav 2017 Jun;11(3):797-807. [doi: 10.1007/s11682-016-9554-1] [Medline: 27180247]
60. Wang W, Ho RLM, Gatto B, Der Veen SMV, Underation MK, Thomas JS, et al. A novel method to understand neural oscillations during full-body reaching: a combined EEG and 3D virtual reality study. IEEE Trans Neural Syst Rehabil Eng 2020 Dec;28(12):3074-3082. [doi: 10.1109/tnsre.2020.3039829]

Abbreviations

ACC: anterior cingulate cortex
BAI: Beck Anxiety Inventory
BFNE: Brief-Fear of Negative Evaluation
DLPFC: dorsolateral prefrontal cortex
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
fMRI: functional magnetic resonance imaging
fNIRS: functional near-infrared spectroscopy
FPPFC: frontopolar prefrontal cortex
HbO_2: oxygenated hemoglobin
HbR: deoxygenated hemoglobin
HC: healthy control
HMD: head-mounted display
ISS: Internalized Shame Scale
K-SAD: Korean Social Avoidance and Distress
K-SIAS: Korean version of the Social Interaction Anxiety Scale
K-SPS: Korean version of the Social Phobia Scale
LSAS: Liebowitz Social Anxiety Scale
MINI: Mini-International Neuropsychiatric Interview
MPFC: medial prefrontal cortex
OFC: orbitofrontal cortex
PCC: posterior cingulate cortex
PERS: Post-Event Rumination Scale
PET: positron emission tomography
PFC: prefrontal cortex
SAD: social anxiety disorder
STAI: State-Trait Anxiety Inventory

VLPFC: ventrolateral prefrontal cortex
VR: virtual reality

Edited by G Eysenbach; submitted 07.07.21; peer-reviewed by JH Baek; comments to author 28.07.21; revised version received 25.08.21; accepted 21.11.21; published 17.12.21.

Please cite as:
Lee H, Choi J, Jung D, Hur JW, Cho CH
The Effects of Virtual Reality Treatment on Prefrontal Cortex Activity in Patients With Social Anxiety Disorder: Participatory and Interactive Virtual Reality Treatment Study
J Med Internet Res 2021;23(12):e31844
URL: https://www.jmir.org/2021/12/e31844
doi:10.2196/31844
PMID:34801979

Original Paper

The Utility of Different Data Standards to Document Adverse Drug Event Symptoms and Diagnoses: Mixed Methods Study

Erina Chan[1,2*], BSc, ACPR; Serena S Small[1,3*], MA; Maeve E Wickham[1,4*], MSc; Vicki Cheng[5*], PharmD; Ellen Balka[1,6*], PhD; Corinne M Hohl[1,3,7*], MD, MHSc

[1]Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada

[2]Vancouver General Hospital Pharmacy Department, Vancouver, BC, Canada

[3]Department of Emergency Medicine, University of British Columbia, Vancouver, BC, Canada

[4]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[5]Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada

[6]School of Communication, Simon Fraser University, Burnaby, BC, Canada

[7]Vancouver General Hospital Emergency Department, Vancouver, BC, Canada

[*]all authors contributed equally

Corresponding Author:
Serena S Small, MA
Centre for Clinical Epidemiology and Evaluation
Vancouver Coastal Health Research Institute
828 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 604 875 4111 ext 55219
Email: Serena.Small@ubc.ca

Abstract

Background: Existing systems to document adverse drug events often use free text data entry, which produces nonstandardized and unstructured data that are prone to misinterpretation. Standardized terminology may improve data quality; however, it is unclear which data standard is most appropriate for documenting adverse drug event symptoms and diagnoses.

Objective: This study aims to compare the utility, strengths, and weaknesses of different data standards for documenting adverse drug event symptoms and diagnoses.

Methods: We performed a mixed methods substudy of a multicenter retrospective chart review. We reviewed the research records of prospectively diagnosed adverse drug events at 5 Canadian hospitals. A total of 2 pharmacy research assistants independently entered the symptoms and diagnoses for the adverse drug events using four standards: Medical Dictionary for Regulatory Activities (MedDRA), Systematized Nomenclature of Medicine (SNOMED) Clinical Terms, SNOMED Adverse Reaction (SNOMED ADR), and International Classification of Diseases (ICD) 11th Revision. Disagreements between research assistants regarding the case-specific utility of data standards were discussed until a consensus was reached. We used consensus ratings to determine the proportion of adverse drug events covered by a data standard and coded and analyzed field notes from the consensus sessions.

Results: We reviewed 573 adverse drug events and found that MedDRA and ICD-11 had excellent coverage of adverse drug event symptoms and diagnoses. MedDRA had the highest number of matches between the research assistants, whereas ICD-11 had the fewest. SNOMED ADR had the lowest proportion of adverse drug event coverage. The research assistants were most likely to encounter terminological challenges with SNOMED ADR and usability challenges with ICD-11, whereas least likely to encounter challenges with MedDRA.

Conclusions: Usability, comprehensiveness, and accuracy are important features of data standards for documenting adverse drug event symptoms and diagnoses. On the basis of our results, we recommend the use of MedDRA.

(J Med Internet Res 2021;23(12):e27188) doi:10.2196/27188

KEYWORDS

adverse drug events; health information technology; data standards

Introduction

Background

Adverse drug events are the harmful and unintended consequences of medication use and are a leading cause of emergency department visits and hospitalizations in Canada and internationally [1-4]. Adverse drug events comprise various types of medication-related problems, including adverse drug reactions (ie, noxious effects that occur within a standard dosing range of a prescription drug). Adverse drug events frequently recur without documentation and communication, which compromises patient safety [5]. The incidence, severity, and recurrence of adverse drug events suggest a need for greater documentation and communication of such events to avoid patients being re-exposed to harmful medications [5].

Adverse drug event reporting is voluntary for clinicians in Canada but has recently become mandatory for hospitals [6]. The implementation of mandatory reporting for hospitals introduces concerns about the added burden of documentation for clinicians. There is often a disconnect between adverse drug event reporting and clinical care activities because of time constraints and a poor fit between standardized nomenclatures built into inflexibly designed reporting systems [7]. The existing electronic medical records include data fields for documenting allergies but can be restrictive and inappropriate for documenting adverse drug reactions and other types of adverse drug events [7]. Furthermore, allergies and adverse drug reactions are a fraction of all the reportable clinically significant adverse drug events [5]. Even when broader input fields are available to document adverse drug events, they are often in free text format, and thus, the resulting data are unstructured, nonstandardized, and prone to misinterpretation. As a result, clinicians who diagnose and treat adverse drug events rarely report them in the existing electronic systems [1,4,8]. Enabling data entry using standardized terminology may reduce the ambiguity of adverse drug event reports, ease the data entry process, improve the utility of the systems, and thereby improve patient safety [9,10]. However, the use of standardized data systems that are incompatible with clinical work may also compromise patient safety and reduce the quality and availability of data for research purposes [11].

Objective

System designers may leverage a number of existing national and international standards to support documentation; however, few studies have examined which data standard is preferable for capturing details about adverse drug events. This study aims to understand and compare the utility of different clinical data standards in capturing adverse drug event symptoms and diagnoses. This has been undertaken in relation to a new law in Canada that mandates the reporting of serious adverse drug reactions [6]. We hope to provide insight into the strengths and weaknesses of different standards as vendors begin to develop software to support adverse drug event documentation.

Methods

Study Design

This was a mixed methods substudy of a multicenter retrospective chart review [5]. We used a convergent mixed methods design in which we collected and analyzed quantitative and qualitative data concurrently and separately and then merged and compared them during the interpretation phase [12].

Setting and Population

We reviewed the research records of all patients who were diagnosed with ≥1 adverse drug event in 1 of the 3 prospective multicenter studies [13-16]. The first study enrolled 1591 patients presenting to the emergency departments of 2 tertiary care hospitals—Vancouver General Hospital (VGH) and St Paul's Hospital—in Vancouver, British Columbia, Canada, from 2008 to 2009 and derived a clinical decision rule to identify the patients at high risk of adverse drug events [13]. The second study enrolled 10,807 patients presenting to the emergency departments of VGH, Lions Gate Hospital (an urban community hospital in North Vancouver, British Columbia), and Richmond Hospital (an urban community hospital in Richmond, British Columbia) between 2011 and 2013 and evaluated the impact of a pharmacist-led medication review on health outcomes [14,16]. The third study enrolled 1529 patients presenting to the emergency departments of VGH, Lions Gate Hospital, and the Ottawa Civic Hospital (an urban tertiary care hospital in Ottawa, Ontario, Canada) from 2014 to 2015 and validated the previously derived clinical decision rule [15].

In all 3 prior studies, the research assistants used a systematic selection algorithm to select and enroll a representative sample of emergency department patients (Multimedia Appendix 1). A clinical pharmacist and physician evaluated all the enrolled patients with adverse drug events at the point of care and documented the events in research and medical records. All the cases in which the clinical pharmacist diagnoses and physician diagnoses were concordant were considered final. An independent committee adjudicated all the cases in which the assessments were discordant or uncertain by reviewing the research and medical records.

Inclusion and Exclusion Criteria

This study included all adverse drug events that met our case definition and were diagnosed in 1 of the 3 primary studies (see *Case Definition* section). We excluded events with alternative diagnoses, those for which records could not be retrieved or were illegible, and those that were not unique with respect to the drug and presenting symptom or diagnosis [13-16].

Case Definition

Adverse drug events included adverse drug reactions, drug interactions, supratherapeutic or subtherapeutic dosing, untreated indications, drug withdrawal, ineffective drugs, nonadherence, and errors in prescribing, dispensing, or medication administration [5,13-16]. These adverse drug events had to be

classified as moderate, resulting in a change in medical management, diagnostic testing, or consulting or severe, resulting in hospital admission, permanent disability, or death [5,16].

Chart Review Data Collection Methods

A total of 2 research assistants (EC, a clinical pharmacist and VC, a pharmacy student) retrospectively reviewed the research records of the enrolled patients. They were independent and blinded to one another's data collection and applied the different data standards to document up to 4 symptoms or diagnoses that they felt were appropriate to describe each adverse drug event using an electronic data collection form (Multimedia Appendix 2). If the research assistants were unable to identify an appropriate symptom or diagnosis, they selected *No Match*. The research assistants then documented whether they thought the terms selected in the data standard accurately described the case.

We conducted a pilot period to ensure the quality of the data collected and identify any potential questions about the application of our research protocol. During the pilot period, the research assistants collected data on a sample of 20 adverse drug events and subsequently provided feedback on the data collection form. We edited the form following the provision of feedback, which the research assistants then piloted on a new sample of 20 records, resulting in a total of 40 records being piloted.

During the pilot period, the research assistants met weekly to discuss discordant cases in which there were disagreements in the identified symptoms or diagnoses for each data standard to ensure consistency in case interpretation. After the pilot period, the research assistants met monthly to discuss the discordant cases in which there was disagreement in the accuracy of the data standard in describing the case. We considered all cases in which the research assistants reached a consensus on the various data standards as final (Multimedia Appendix 3).

We randomly selected 100 adverse drug events for the research assistants to assess twice to evaluate intrarater reliability.

Qualitative Data Collection Methods

During the chart review, the research assistants electronically recorded notes on their process, general impressions, and any case-specific challenges they encountered for each data standard. A qualitative researcher (SS) attended the meetings between the research assistants to observe the discussion of the discordant cases and took notes on the discussion to capture emerging themes and points of convergence and divergence. This produced a richer understanding of the human factors that influence the perceived utility of data standards.

Data Standards

We used four data standards to document the symptoms and diagnoses of each adverse drug event: Systematized Nomenclature of Medicine (SNOMED) Health Concern and Diagnosis (SNOMED HC) reference set, SNOMED Adverse Reaction (SNOMED ADR) reference set, Medical Dictionary for Regulatory Activities (MedDRA), and International Classification of Diseases (ICD) 11th Revision. We selected these as various levels of government and other organizations in Canada recommend their use in different clinical contexts related to adverse drug event reporting and documentation.

SNOMED HC Reference Set

SNOMED Clinical Terms (SNOMED CT) is an international clinical terminology coding system that includes diagnoses, signs, symptoms, and diagnostic procedures. We used the SNOMED CT Canadian Edition, which was developed specifically for use in Canada and released in October 2018 [17]. The SNOMED HC reference set is a subset of SNOMED CT, which is designed to map terminology to the ICD-9 and ICD-10 codes and the Canadian Emergency Department Diagnosis Shortlist for use in electronic medical records and clinical information systems. We included SNOMED CT as it is maintained and recommended by Canada Health Infoway to support the capture and exchange of clinical data in Canada [18].

SNOMED ADR Reference Set

The SNOMED ADR reference set highlights the allergies and intolerances found in SNOMED CT. This is a baseline reference set under development, which will continue to expand based on feedback. The Northern Health Authority in British Columbia has integrated this data standard into their electronic medical record system. Users select the terminology through a search function, which creates a filtered dropdown selection list. We included SNOMED ADR as it is under development for use specifically in British Columbia [19].

MedDRA Preferred Terms

MedDRA version 22.0 is an international standardized medical terminology dictionary that supports classification of adverse event information associated with biopharmaceuticals and other medical products [20]. MedDRA's hierarchical structure comprises 5 terminological levels that map to one another, arranged from very specific to very general. We used the *Preferred Terms* level, which presents terms as distinct descriptors for symptoms, signs, disease diagnoses, therapeutic indications, investigations, surgical or medical procedures, and medical, social, or family history characteristics. We modified some terms from their original British spelling to their American spelling by referring to corresponding Lower Level Terms in the MedDRA hierarchy. We included MedDRA as its use is recommended by Health Canada for adverse reaction reports submitted to their pharmacovigilance database [21].

International Classification of Diseases 11th Revision

ICD-11 is an international standard for reporting diseases and health conditions for clinical and research purposes [22]. The World Health Organization released the 11th revision in 2018 for piloting, and it will come into use in 2022. This version provides a coding system designed for easy adoption into electronic environments and updated clinical content, including explicit coding to capture adverse drug events. We included ICD-11 as it is used by physicians in British Columbia for claims submissions to the provincial Medical Services Plan [23].

Quantitative Analysis

We used descriptive statistics to describe the baseline characteristics of all included adverse drug events using

proportions. To determine the coverage of a specific data standard to capture an event, we used the research assistants' ratings of whether a data standard contained a match for the characteristics of an adverse drug event. We used the consensus assessments to calculate the frequency and proportion of adverse drug events with symptoms or diagnoses found within a given data standard.

We allowed entries for up to 4 terms per data standard. To identify whether the research assistants agreed on the term selected within a given data standard, we examined whether the first selected term for each data standard matched (first term match). Then, we examined whether there were any matches across the 4 terms between the terms used by both research assistants for each data standard (all terms match).

Qualitative Analysis

We coded comment fields from the data collection forms and notes from our observations using NVivo (version 12; QSR International) qualitative data analysis software. We began by inductively coding field notes to generate a provisional coding structure that we then applied to the comment fields. We iteratively reviewed the data and coding to identify emerging themes. We completed an interim review of the coding and emerging findings to contextualize the results with quantitative data and validate them with the research assistants' experiences. Following discussion, we generated a final coding structure and used a descriptive approach to describe the classification challenges.

Results

Quantitative Results

Figure 1 displays the flow of patients in the study sample. Overall, we included 673 adverse drug events in 573 patients in our sample. The top 5 most common culprit drugs and diagnoses of the included adverse drug events are presented in Table 1. The most common culprit medication overall was warfarin (62/673, 9.2%, 95% CI 7.1%-11.7%), and the most common diagnosis was allergic reaction (38/673, 5.7%, 95% CI 4%-7.7%).

Table 2 displays the coverage of the different data standards for the adverse drug events in our sample. Of the data standards, MedDRA (671/673, 99.7%, 95% CI 98.9%-99.9%) and ICD-11 (667/673, 99.1%, 95% CI 98%-99.6%) had the highest frequency of having an appropriate symptom or diagnosis available. SNOMED ADR had the lowest frequency (576/672, 85.7%, 95% CI 82.7%-88%).

Figure 1. Flow diagram of patients through the study. ED: emergency department; ADE: adverse drug event.

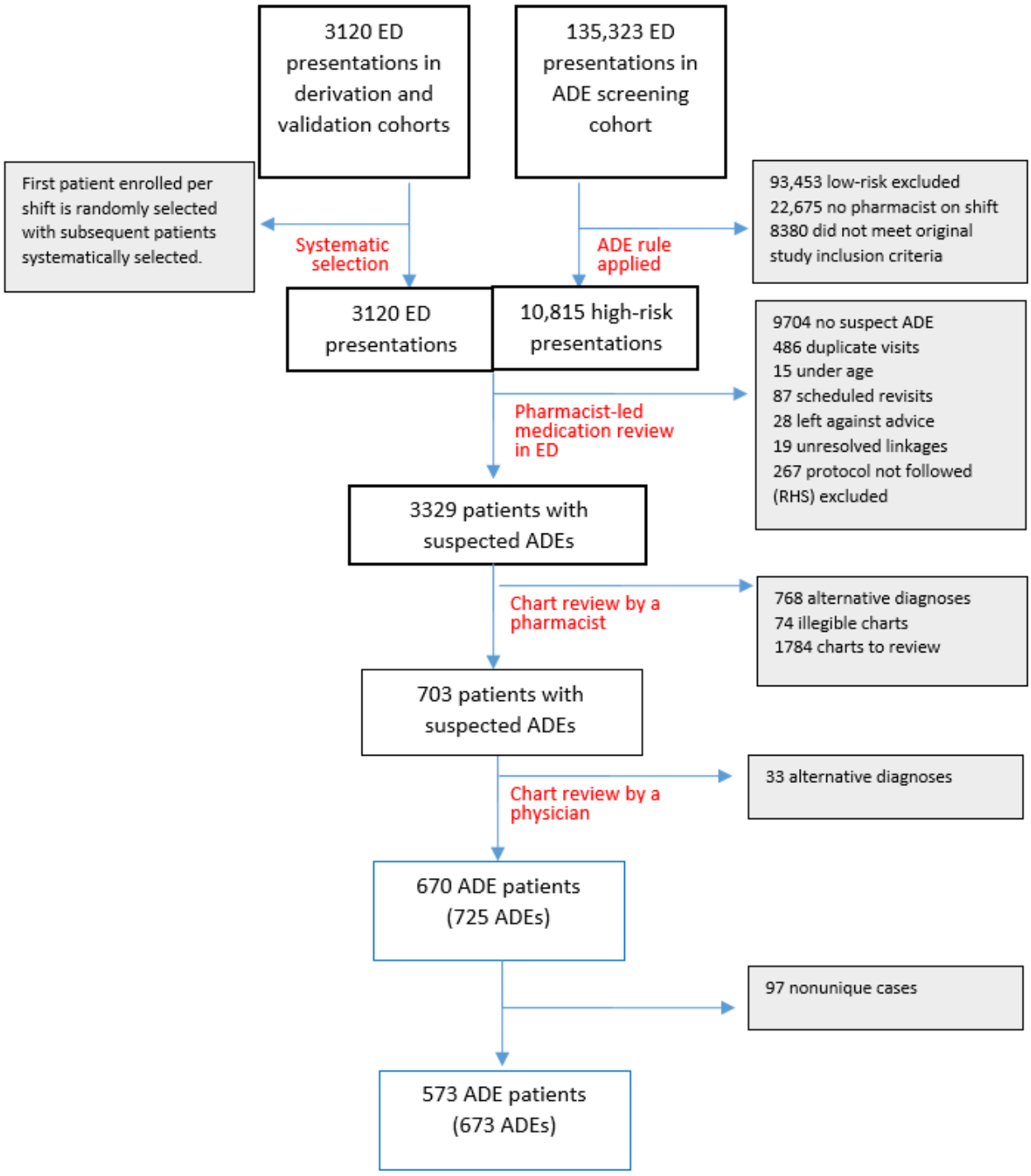


Table 1. Adverse drug event (ADE) characteristics of the study sample (N=673).

Characteristics	ADEs, n (%)
Culprit medication	
Warfarin	62 (9.2)
Furosemide	37 (5.5)
Acetylsalicylic acid	37 (5.5)
Hydrochlorothiazide	26 (3.9)
Insulin	21 (3.1)
ADE diagnosis	
Allergic reaction	38 (5.7)
Laboratory abnormality	36 (5.4)
Stroke	22 (3.3)
Hypoglycemia	20 (3)
Atrial fibrillation	19 (2.8)

Table 2. The percentage of having an appropriate symptom or diagnosis available to describe the adverse drug event cases by data standard (N=673).

Data standard	Symptom or diagnosis option available, n (%, 95% CI)
MedDRA[a]	671 (99.7, 98.9-99.9)
SNOMED ADR[b,c]	576 (85.7, 82.7-88)
SNOMED HC[d]	650 (96.6, 94.9-97.79)
ICD-11[e]	667 (99.1, 98-99.6)

[a]MedDRA: Medical Dictionary for Regulatory Activities.

[b]SNOMED ADR: Systematized Nomenclature of Medicine Adverse Reaction.

[c]N=672.

[d]SNOMED HC: Systematized Nomenclature of Medicine Health Concern and Diagnosis.

[e]ICD-11: International Classification of Diseases 11th Revision.

Table 3 presents the percentage agreement between the research assistants for the first term match and for any term match for each data standard. For the first term match, SNOMED HC (409/673, 60.8%, 95% CI 57%-64%) yielded the most matches between the research assistants and ICD-11 (286/673, 42.5%, 95% CI 38.8%-46.3%) yielded the fewest matches. In terms of having any term match, MedDRA performed the best (673/673, 100%, 95% CI 99.4%-100%) and ICD-11 yielded the lowest proportion of matches (583/673, 86.6%, 95% CI 83.9%-89%). Semantic differences between terms with identical meanings within a data standard may have artificially lowered the number of matches for a given data standard. For example, in SNOMED HC, 1 research assistant selected the term *frank hematuria*, whereas the second research assistant selected the term *blood in urine* to describe hematuria. In SNOMED ADR, 1 research assistant selected *muscle weakness*, and the second selected *asthenia* to describe weakness. In ICD-11, 1 research assistant selected *candidiasis of lips or oral mucous membranes*, and the second selected *thrush disorder* to describe thrush. The complete list of adverse drug events that did not have a match on any of the data standard terms is presented in Multimedia Appendix 4.

Table 3. The percentage agreement between the research assistants in coding the adverse drug events by data standards (N=673).

Data standard	First term match, n (%)	Any terms match, n (%)
MedDRA[a]	400 (59.4)	673 (100)
SNOMED ADR[b]	379 (56.3)	601 (89.3)
SNOMED HC[c]	409 (60.8)	629 (93.5)
ICD-11[d]	286 (42.5)	583 (86.6)

[a]MedDRA: Medical Dictionary for Regulatory Activities.

[b]SNOMED ADR: Systematized Nomenclature of Medicine Adverse Reaction.

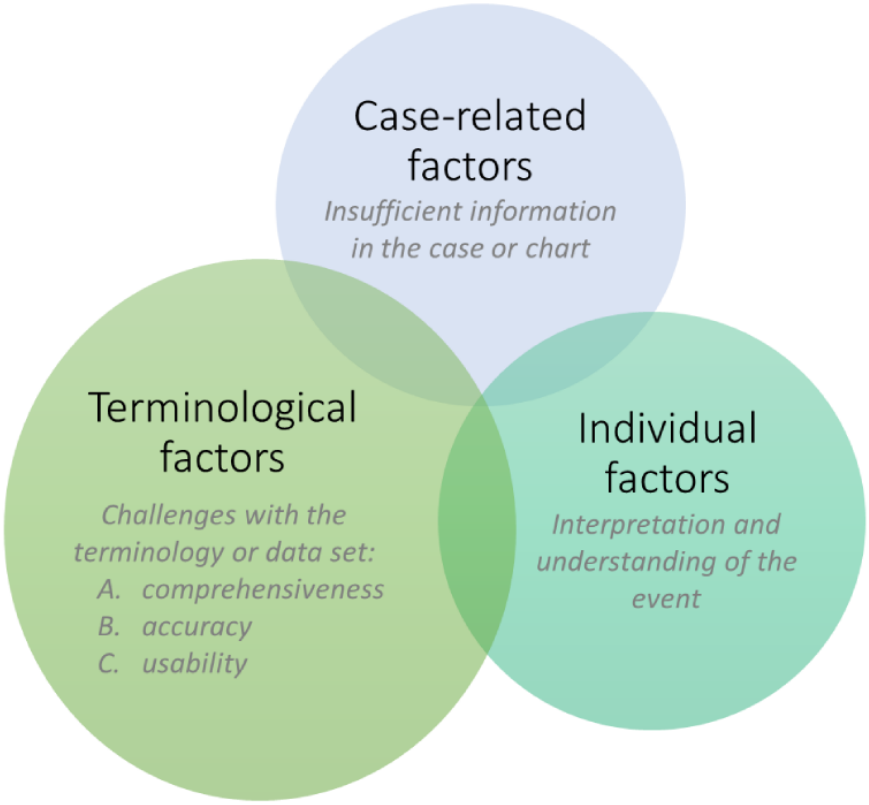
[c]SNOMED HC: Systematized Nomenclature of Medicine Health Concern and Diagnosis.

[d]ICD-11: International Classification of Diseases 11th Revision.

Qualitative Results

We found 3 primary factors that affected the classification of adverse drug event symptoms and diagnoses: (1) terminological factors specific to the terminology or data standard, (2) case-related factors in which there was not enough information in the patient's chart to classify the event appropriately, and (3) individual factors related to interpretation and recall (Figure 2).

Figure 2. Classification challenges in qualitative analysis.



Terminological Factors

Terminological challenges were the most common factors that affected the adverse drug event classification. We assessed the overall utility of each data standard according to 3 key terminological factors: comprehensiveness (conceptual coverage or breadth; eg, is the data source comprehensive enough to select appropriate terms?), accuracy (terminological correctness or exactness; eg, does the available terminology accurately describe the case?), and usability (ease of use; eg, is it easy to find an appropriate term using the data source?).

We used the ability to locate any term to describe the main symptom or diagnosis as a proxy variable for comprehensiveness. For example, both research assistants noted that the primary symptom and diagnosis of an adverse drug event were *hemiparesis* and *stroke*; however, there was no option to describe this case in SNOMED ADR. Poor comprehensiveness was found most often when using SNOMED ADR and rarely when using MedDRA, which is consistent with our quantitative findings related to coverage.

Issues with the accuracy of data sources emerged when the terms did not fully capture or represent the case, were too

specific, or were too vague. In the case of the main adverse drug event symptom or diagnosis being *fall*, the research assistant found partial terminology in SNOMED ADR, including *weakness* and *syncope and collapse*; however, during consensus, the research assistant noted that the patient did not have syncope. In this case, although some terms were available, they did not produce a complete and clinically meaningful or accurate description of the event. An instance in which the terminology was too specific arose with ICD-11, wherein the only term available to describe a hematoma case included the qualifier *of other specified site complicating a procedure*; however, there was no indication that this hematoma was in fact complicating a procedure. An example of exceedingly broad or vague terminology arose when the research assistants could not find a specific term for *Clostridium difficile* in SNOMED ADR, which led to a discussion of whether the term *diarrhea* was adequate. These challenges occurred most often with SNOMED ADR and least often with MedDRA.

Unusual terminology and phrasing or unfamiliar spelling compromised the usability of the data sources. Usability challenges emerged when the research assistants reported that they felt a term adequately described the case but that it was difficult to find, it was only identified during consensus, or they had to rely on external sources to identify the term (eg, Google). In the case of a *rash*, the research assistants identified the term *allergic disorder of the skin* in ICD-11 as the closest descriptor but felt that it was atypical phrasing. Another issue was the use of British English spellings for certain terms (eg, *haemorrhage* instead of *hemorrhage*). Usability challenges were most common with ICD-11; however, the research assistants encountered them less often than the issues with accuracy or comprehensiveness.

Case-Related Factors

In some cases, there was insufficient information in the chart to classify the event independent of the data source. For example, in a case, the research assistant noted that the selected term was broad enough, but the classification would have been improved if there were more details about the patient's documented *bizarre behavior*. The research assistants also encountered cases with insufficient information to classify the event in the context of the data standard's limitations, often because of vague case descriptions that could only be classified using high specificity terms. For example, in SNOMED ADR, the terms to describe headaches were often too specific, such as *frontal headache* or *migraine with aura*, whereas case descriptions tended to use only the term *headache*.

Individual Factors

Individual factors, such as recall and interpretation, had an effect on the classification of events. During consensus, the research assistants discussed instances where 1 research assistant did not identify the correct term that the other had identified. This occurred because they forgot the terminology (eg, the research assistant had been searching for the term *kidney* rather than *renal* to describe abnormal renal function), did not consider alternate wording (eg, the research assistant did not think of the term *spasticity* to refer to *rigidity* or *stiffness*), or were unable to locate a term that they felt was acceptable (eg, the research assistant could not find a term related to the patient's history of noncompliance). In almost all of these instances, the research assistant agreed with the other's selection during the consensus.

Discussion

Principal Findings

Previous studies have demonstrated gaps in the existing terminological standards in health care [24,25]. Our findings add a nuanced examination of these gaps and other shortcomings of multiple terminological standards rooted in clinical practice. We explored the utility of 4 data standards to document adverse drug event symptoms and diagnoses. Our quantitative analysis demonstrated that MedDRA and ICD-11 were most likely to have an appropriate symptom or diagnosis available. MedDRA most often had any match documented, whereas ICD-11 had the fewest matches. SNOMED HC performed the best in terms of the first term that the research assistants selected for matching. SNOMED ADR performed the worst in terms of having the lowest capture of a symptom or diagnosis. These results are consistent with our qualitative findings. The research assistants were least likely to encounter terminological challenges with MedDRA and most likely with SNOMED ADR. We found that ICD-11 was most likely to present usability challenges because of unusual terminology or spelling, which may provide a rationale for why ICD-11 had the fewest matches. The research assistants also found ICD-11 to be the most time consuming for searching terms because of the lengthy list of returned matches with descriptive terms, whereas SNOMED HC and SNOMED ADR were the least time consuming, with a shorter list of returned matches with more straightforward terms to select from. Overall, across all the indicators, we found that MedDRA was the strongest data standard, whereas SNOMED ADR performed poorest. We acknowledge that SNOMED ADR is a working data set at this time and thus could be strengthened through further study and use.

Implementing clinical information systems with data standards that lack comprehensiveness, accuracy, or usability in clinical practice will affect data entry and generate downstream negative effects on data quality and the information generated. In the absence of correct or accurate terminology, research assistants were more likely to make compromises or use workarounds by selecting a term that was close enough or only partially described the event. In clinical practice, challenges with data entry, along with time constraints and other external pressures, may result in a clinician opting to abandon data entry altogether, thus lowering data quantity and quality. Conversely, semantic standardization may lead to more consistent and complete reporting for pharmacovigilance activities if the appropriate data standard is used, which may produce higher data quality and facilitate data analysis [26-28]. Reliable coding for adverse drug reactions is likely to yield more meaningful data for the end user and may facilitate data integration across different electronic health systems [28,29].

Recent efforts to map terminology across different standards may be used to develop clinical information systems with specific data standards while facilitating data integration across systems and pharmacovigilance organizations. Reich et al [30]

demonstrated that it is feasible to map ICD-9 diagnosis codes for medical conditions to SNOMED CT and MedDRA, making both suitable options for standard vocabularies. However, to our knowledge, there has been no study that has compared these data sets with one another or investigated their use specifically for documenting adverse drug events [30]. The WEB-RADR 2 Project seeks to develop a bidirectional mapping of a subset of pharmacovigilance terms between SNOMED CT and MedDRA [31]. Mapping and testing were scheduled to be completed in 2020, with a production version of the map available to SNOMED and MedDRA users in 2021. In addition, the National Institutes of Health's National Library of Medicine has also developed a Unified Medical Language System that maps SNOMED CT to ICD-10 to support reimbursement and statistical analyses [32]. Further research should examine the effect of mapping data standards on data quality.

In addition to the efforts to map terminology across data standards, advances in natural language processing increasingly offer a new and promising approach to the analysis of adverse drug event reports for pharmacovigilance activities. Using natural language processing, system designers may enable free text data entry from clinicians to increase the ease of use. Such systems would then algorithmically analyze the entered data to produce standardized data for monitoring and regulatory purposes [33]. A recent systematic review found that many studies on natural language processing of incidents, adverse events, and medical error reports have focused primarily on binary classification, which does not account for the complexity of adverse drug event documentation that we sought to capture and limits the subsequent clinical utility of data to support continuity of care [33]. In addition, in instances where cases contain insufficient information for classification, which we encountered in this study, natural language processing is unlikely to improve the results. Continued research in this field should explore natural language processing, aim to produce multimodal analyses of reports, and increase integration across clinical information systems.

In the absence of an agreed-upon standard for data capture and with the advent of increased mapping across terminologies, we suggest that health system designers prioritize the implementation of a data standard that is clinically useful and relevant to ensure high usability for clinicians who are asked to document the event while being immersed in clinical activities. For this purpose, MedDRA was the strongest data standard among the data sets in our study, and we recommend it be used as the standard for Canadian pharmacovigilance activities in support of federal legislation that requires all Canadian health institutions to report serious adverse drug reactions to Health Canada [6]. MedDRA is also currently used in other pharmacovigilance systems, such as the US Food and Drug Administration Adverse Drug Event Reporting System and Vaccine Adverse Event Reporting System databases, the European Medicines Agency Eudrawatch system, and the Japanese prescription event monitoring system, which makes it a strong option to advance international collaborative efforts in pharmacovigilance.

Limitations

There are limitations to this study. Our research team is more familiar with MedDRA, which may have led to bias when selecting the most comprehensive data standard. However, one of our research assistants had previously never worked with our team and thus was unfamiliar with MedDRA at the outset of this study. We also relied on the American translation of British MedDRA terminology, which may have facilitated the identification of terms and resulted in more matches. The ICD-11 terms remained in their original British spelling, which may have initially resulted in fewer matches. We observed that this effect was offset as the research assistants became familiar with these patterns over time and with use. SNOMED ADR is designed to describe adverse drug reactions; however, we applied it to describe a broader range of adverse drug events. This may have positioned SNOMED ADR to perform poorly from the outset and falsely lowered its capture of a symptom or diagnosis compared with the other data standards. We were also unable to obtain more information than what was available in the research records, as they were from a previous multicenter chart review study, which may have limited the terms we could have selected for a given data standard.

Conclusions

Usability, comprehensiveness, and accuracy are the key features of a data standard for documenting adverse drug event symptoms and diagnoses. On the basis of these factors, we found that MedDRA is the most suitable data standard for coding adverse drug events in electronic reporting systems. Although data standardization is important, not all standards are created equally. As our analyses demonstrate, each data standard has different affordances and constraints. Hence, it is important to critically evaluate competing standards to ensure that the data standards adopted in clinical information systems support patient safety rather than compromise it. When the appropriate data standard is selected, the standardized terminology may result in more consistent adverse drug event documentation and better data quality and quantity as a by-product of routine care.

Acknowledgments

The authors acknowledge and thank Amber Cragg for her support in the initial project coordination and protocol development. This research was sponsored by the Canadian Institutes of Health Research (FDN-148475).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient enrollment algorithm.
[DOCX File , 50 KB - jmir_v23i12e27188_app1.docx]

Multimedia Appendix 2
Data collection form.
[DOCX File , 25 KB - jmir_v23i12e27188_app2.docx]

Multimedia Appendix 3
Consensus data collection form.
[DOCX File , 16 KB - jmir_v23i12e27188_app3.docx]

Multimedia Appendix 4
Adverse drug events with no matching terms.
[DOCX File , 49 KB - jmir_v23i12e27188_app4.docx]

References

1. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med 2004 May 18;140(10):795-801. [doi: 10.7326/0003-4819-140-10-200405180-00009] [Medline: 15148066]
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. J Am Med Assoc 1998 Apr 15;279(15):1200-1205. [doi: 10.1001/jama.279.15.1200] [Medline: 9555760]
3. Budnitz DS, Lovegrove MC, Shehab N, Richards CL. Emergency hospitalizations for adverse drug events in older Americans. N Engl J Med 2011 Nov 24;365(21):2002-2012. [doi: 10.1056/nejmsa1103053]
4. Zed PJ, Abu-Laban RB, Balen RM, Loewen PS, Hohl CM, Brubacher JR, et al. Incidence, severity and preventability of medication-related visits to the emergency department: a prospective study. Can Med Asso J 2008 Jun 03;178(12):1563-1569 [FREE Full text] [doi: 10.1503/cmaj.071594] [Medline: 18519904]
5. Hohl CM, Woo SA, Cragg A, Wickham ME, Ackerley C, Scheuermeyer F, et al. Repeat adverse drug events associated with outpatient medications: a descriptive analysis of 3 observational studies in British Columbia, Canada. CMAJ Open 2019 Jul 18;7(3):446-453 [FREE Full text] [doi: 10.9778/cmajo.20180190] [Medline: 31320328]
6. Protecting Canadians from Unsafe Drugs Act (Vanessa's Law) amendments to the Food and Drugs Act (Bill C-17). Government of Canada - Health Canada. 2015. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/legislation-guidelines/protecting-canadians-unsafe-drugs-act-vanessa-law-amendments-food-drugs-act.html [accessed 2021-11-27]
7. Peddie D, Small SS, Badke K, Bailey C, Balka E, Hohl CM. Adverse drug event reporting from clinical care: mixed-methods analysis for a minimum required dataset. JMIR Med Inform 2018 Jun 28;6(2):e10248 [FREE Full text] [doi: 10.2196/10248] [Medline: 29954724]
8. Hohl CM, Small SS, Peddie D, Badke K, Bailey C, Balka E. Why clinicians don't report adverse drug events: qualitative study. JMIR Public Health Surveill 2018 Feb 27;4(1):e21 [FREE Full text] [doi: 10.2196/publichealth.9282] [Medline: 29487041]
9. Molokhia M, Tanna S, Bell D. Improving reporting of adverse drug reactions: systematic review. Clin Epidemiol 2009 Aug 09;1:75-92 [FREE Full text] [doi: 10.2147/clep.s4775] [Medline: 20865089]
10. van der Linden CM, Jansen PA, van Marum RJ, Grouls RJ, Korsten EH, Egberts AC. Recurrence of adverse drug reactions following inappropriate re-prescription: better documentation, availability of information and monitoring are needed. Drug Saf 2010 Jul 01;33(7):535-538. [doi: 10.2165/11532350-000000000-00000] [Medline: 20553055]
11. Tolar M, Balka E. Beyond individual patient care: enhanced use of EMR data in a primary care setting. Stud Health Technol Inform 2011;164:143-147. [Medline: 21335702]
12. Creswell JW, Clark VL. Designing and Conducting Mixed Methods Research. 3rd ed. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
13. Hohl CM, Yu E, Hunte GS, Brubacher JR, Hosseini F, Argent CP, et al. Clinical decision rules to improve the detection of adverse drug events in emergency department patients. Acad Emerg Med 2012 Jun 11;19(6):640-649 [FREE Full text] [doi: 10.1111/j.1553-2712.2012.01379.x] [Medline: 22687179]
14. Hohl CM, Partovi N, Ghement I, Wickham ME, McGrail K, Reddekopp LN, et al. Impact of early in-hospital medication review by clinical pharmacists on health services utilization. PLoS One 2017 Feb 13;12(2):e0170495 [FREE Full text] [doi: 10.1371/journal.pone.0170495] [Medline: 28192477]
15. Hohl CM, Badke K, Zhao A, Wickham ME, Woo SA, Sivilotti ML, et al. Prospective validation of clinical criteria to identify emergency department patients at high risk for adverse drug events. Acad Emerg Med 2018 Sep 11;25(9):1015-1026 [FREE Full text] [doi: 10.1111/acem.13407] [Medline: 29517818]

16. Hohl CM, McGrail K, Sobolev B. The effect of pharmacist-led medication review in high-risk patients in the emergency department: an evaluation protocol. CMAJ Open 2015 Mar 26;3(1):103-110 [FREE Full text] [doi: 10.9778/cmajo.20140010] [Medline: 25844362]
17. 5-Step briefing. SNOMED International. 2020. URL: http://www.snomed.org/snomed-ct/five-step-briefing [accessed 2021-11-30]
18. SNOMED CT CA/ SNOMED CT. Canada Health Infoway. URL: https://infocentral.infoway-inforoute.ca/en/standards/canadian/snomed-ct [accessed 2021-07-15]
19. Health information and interoperability standards catalogue. Government of British Columbia. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/health-information-standards/standards-catalogue [accessed 2021-07-15]
20. MedDRA Hierarchy. MedDRA. URL: https://www.meddra.org/how-to-use/basics/hierarchy [accessed 2021-07-15]
21. Reporting adverse reactions to marketed health products: guidance document for industry. Health Canada. 2018. URL: https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/reports-publications/medeffect-canada/reporting-adverse-reactions-marketed-health-products-guidance-industry/reporting-adverse-reactions-marketed-health-products-guidance-industry.pdf [accessed 2021-07-15]
22. International statistical classification of diseases and related health problems (ICD). World Health Organization. URL: https://www.who.int/classifications/icd/en/ [accessed 2021-07-15]
23. Diagnostic code descriptions (ICD-9). Government of British Columbia. URL: https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/msp/physicians/diagnostic-code-descriptions-icd-9 [accessed 2021-07-15]
24. Hammond WE. The making and adoption of health data standards. Health Aff (Millwood) 2005 Sep;24(5):1205-1213. [doi: 10.1377/hlthaff.24.5.1205] [Medline: 16162564]
25. Harris MR, Langford LH, Miller H, Hook M, Dykes PC, Matney SA. Harmonizing and extending standards from a domain-specific and bottom-up approach: an example from development through use in clinical applications. J Am Med Inform Assoc 2015 May;22(3):545-552. [doi: 10.1093/jamia/ocu020] [Medline: 25670750]
26. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical data reuse or secondary use: current status and potential future progress. Yearb Med Inform 2017 Aug 19;26(01):38-52. [doi: 10.1055/s-0037-1606528]
27. Elkin PL, Trusko BE, Koppel R, Speroff T, Mohrer D, Sakji S, et al. Secondary use of clinical data. Stud Health Technol Inform 2010;155:14-29. [Medline: 20543306]
28. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. P T 2018 Jun;43(6):340-351 [FREE Full text] [Medline: 29896033]
29. Bousquet C, Lagier G, Louët A, Le Beller C, Venot A, Jaulent M. Appraisal of the MedDRA conceptual structure for describing and grouping adverse drug reactions. Drug Saf 2005;28(1):19-34. [doi: 10.2165/00002018-200528010-00002] [Medline: 15649103]
30. Reich C, Ryan PB, Stang PE, Rocca M. Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases. J Biomed Inform 2012 Aug;45(4):689-696 [FREE Full text] [doi: 10.1016/j.jbi.2012.05.002] [Medline: 22683994]
31. SNOMED CT - MedDRA mapping. European Industry MedDRA User Group Webinar. 2019. URL: https://web-radr.eu/wp-content/uploads/2019/10/european-industry-meddra-user-group-webinar-agenda-december-2019.pdf [accessed 2021-11-30]
32. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):267-270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
33. Young IJ, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019 Dec;132:103971. [doi: 10.1016/j.ijmedinf.2019.103971] [Medline: 31630063]

Abbreviations

ICD: International Classification of Diseases
MedDRA: Medical Dictionary for Regulatory Activities
SNOMED: Systematized Nomenclature of Medicine
SNOMED ADR: Systematized Nomenclature of Medicine Adverse Reaction
SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms
SNOMED HC: Systematized Nomenclature of Medicine Health Concern and Diagnosis
VGH: Vancouver General Hospital

Edited by R Kukafka; submitted 15.01.21; peer-reviewed by S Hume, K Fuji; comments to author 28.06.21; revised version received 16.08.21; accepted 10.11.21; published 10.12.21.

Please cite as:
Chan E, Small SS, Wickham ME, Cheng V, Balka E, Hohl CM
The Utility of Different Data Standards to Document Adverse Drug Event Symptoms and Diagnoses: Mixed Methods Study
J Med Internet Res 2021;23(12):e27188
URL: https://www.jmir.org/2021/12/e27188
doi:10.2196/27188
PMID:34890351

Original Paper

An Integrated Model to Improve Medication Reconciliation in Oncology: Prospective Interventional Study

Alessandro Passardi[1], MD; Patrizia Serra[2], BSc; Caterina Donati[3], BPharm; Federica Fiori[4], RN; Sabrina Prati[4], RN; Roberto Vespignani[5], BEng; Gabriele Taglioni[6], BPharm; Patrizia Farfaneti Ghetti[7], BPharm; Giovanni Martinelli[8], MD; Oriana Nanni[2], MStat; Mattia Altini[9], MD; Giovanni Luca Frassineti[1], MD; Martina Vittoria Minguzzi[10], MPharm

[1]Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[2]Unit of Biostatistics and Clinical Trials, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[3]Oncology Pharmacy Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[4]Oncology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[5]IT Service, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[6]A.S.SO.FARM Community Pharmacy, Ravenna, Italy
[7]Federfarma Community Pharmacy, Rimini, Italy
[8]Hematology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy
[9]Healthcare Administration, Azienda Unità Sanitaria Locale (AUSL) Romagna, Ravenna, Italy
[10]Healthcare Administration, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy

Corresponding Author:
Roberto Vespignani, BEng
IT Service
IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori"
Via P. Maroncelli 40
Meldola, 47014
Italy
Phone: 39 0543 739992
Fax: 39 0543 739151
Email: roberto.vespignani@irst.emr.it

Abstract

Background: Accurate medication reconciliation reduces the risk of drug incompatibilities and adverse events that can occur during transitions in care. Community pharmacies (CPs) are a crucial part of the health care system and could be involved in collecting essential information on conventional and supplementary drugs used at home.

Objective: The aim of this paper was to establish an alliance between our cancer institute, Istituto Romagnolo per lo Studio dei Tumori (IRST), and CPs, the latter entrusted with the completion of a pharmacological recognition survey. We also aimed to integrate the national information technology (IT) platform of CPs with the electronic medical records of IRST.

Methods: Cancer patients undergoing antiblastic treatments were invited to select a CP taking part in the study and to complete the pharmacological recognition step. The information collected by the pharmacist was sent to the electronic medical records of IRST through the new IT platform, after which the oncologist performed the reconciliation process.

Results: A total of 66 CPs completed surveys for 134 patients. An average of 5.9 drugs per patient was used at home, with 12 or more used in the most advanced age groups. Moreover, 60% (80/134) of the patients used nonconventional products or critical foods. Some potential interactions between nonconventional medications and cancer treatments were reported.

Conclusions: In the PROF-1 (Progetto di Rete in Oncologia con le Farmacie di comunità della Romagna) study, an alliance was created between our cancer center and CPs to improve medication reconciliation, and a new integrated IT platform was validated.

Trial Registration: ClinicalTrials.gov NCT04796142; https://clinicaltrials.gov/ct2/show/NCT04796142

(J Med Internet Res 2021;23(12):e31321) doi:10.2196/31321

KEYWORDS

medication recognition; medication reconciliation; IT platform; community pharmacies; healthcare transitions; pharmacy; oncology; drug incompatibility; information technology; drug interactions

Introduction

Medication reconciliation is the process of drawing up a complete and accurate list of all medications being taken by an individual patient, including drug name, dosage, frequency, and route of administration, and comparing them with the medications listed in the patient's medical records or medication prescriptions. The aim of reconciliation is to reduce the risk of errors of omission, duplication, incorrect doses or timing, and adverse drug-drug or drug-disease interactions [1-3]. The Joint Commission on Accreditation of Health Care Organizations added the concept of reconciliation across the continuum of care as a national patient safety goal [4]. It has also been defined as one of the best strategies for maintaining the quality of care by the Agency for Healthcare Research and Quality [5] and is one of the 5 elements in the "High 5s Project" launched by the World Health Organization [6].

Reconciliation must be periodically performed at both the hospital and territorial level [7] and at every transition of care [8], especially when new medications are prescribed or rewritten as several professionals may be involved. The efficacy and quality of this process depends mainly on a preliminary pharmacological survey (recognition step), which not only takes into account the medications prescribed to the patient, but also the phytotherapeutic and homeopathic products, supplements, and foods taken, as well as any habit that might negatively affect patient safety [5]. The success of the recognition step depends on the interlocutor's ability to promote within the patient a sense of empowerment and responsibility in relation to treatment adherence [9]. The use of innovative drugs for the treatment of cancer is constantly increasing [10], and many are unknown to the vast majority of nonexperts in the field. This increases the risk of pharmacological discrepancies in the medications prescribed (or self-prescribed) in the interval between each access to the cancer center caring for the patient [11-13]. In particular, nonconventional medicinal and health preparations such as over-the-counter products, herbal medicines, and supplements are sources of potential pharmacological discrepancy and should be closely monitored [5]. Cancer patients frequently resort to such remedies in an attempt to improve their quality of life, often compromised by the side effects of the treatments, or to help them cope with the emotional aspects related to living with cancer.

Therapy errors have economic implications. The World Health Organization has estimated the cost of therapy errors to be at US $42 billion annually and has set a goal of reducing these costs (which are attributed to weaknesses in health care systems) by 50% within the next 5 years [14]. In the area of oncology, the extent of the economic risk is increased by the burden of expected outcomes and by the high costs of anticancer treatments for the UK National Health Service. In 2018, in Italy, these costs were estimated to be at €5659 million (US $6360 million), constituting the first item of public health expenditure and exceeding that of cardiovascular drugs [15].

Community pharmacies (CPs) are a crucial part of the health care system. They are neighborhood health care facilities whose activities include dispensing medications, treating minor ailments, and offering advice on well-being. CP staff also frequently build up a close relationship with their clients, and they are highly knowledgeable about food supplements as they are responsible for around 80% of the sales of these products on the market [16]. It is therefore relatively easy for CPs to gather information to obtain a complete and accurate picture of the use of conventional drugs and supplements at home.

The territory of Romagna has a population of 1,281,243, spread over 3 provinces (Forlì-Cesena, Ravenna, and Rimini), and represents 28.6% (366,436/1,281,243) of the Emilia-Romagna region in northern Italy. There are 356 CPs in Romagna. Our cancer institute (IRCCS Istituto Romagnolo per lo Studio dei Tumori [IRST], "Dino Amadori") is also based in this area. The center has a high level of computerization and standardization of all therapeutic, clinical, and experimental processes, guaranteeing maximum safety for patients and full traceability of the actions carried out by each professional, clinician, pharmacist, and nurse during their daily activities. The national association representing private CPs (Federfarma) recently developed an information technology (IT) platform (Dottorfarma) in collaboration with an e-commerce company (Promofarma) that can be used by CPs nationwide.

Given these premises, an alliance created between IRST and the CPs of Romagna seemed to provide the perfect opportunity [17] to improve the pharmacological reconciliation process and to bridge the "pharmaceutical gap" between the health care system and the patients. We carried out a prospective, interventional, nonpharmacological study of the first phase of the network project in oncology with the community pharmacies of Romagna (Progetto di Rete in Oncologia con le Farmacie di comunità della Romagna [PROF-1]) on a new model for medication reconciliation. Our overall aims were to create an organizational model and IT platform, assess their ability to promote the shared management of therapies by patients, pharmacists, and health care professionals (thus improving adherence to treatment and the home management of side effects), and evaluate the acceptability of this model by patients.

Methods

The team of investigators included the IRST oncology group (oncologists, hospital pharmacists, and nurses), IRST and Federfarma IT Services, CPs, and scientific representatives of private and public pharmacy associations. The study team also worked closely with lawyers from Federfarma and Promofarma to take care of the delicate aspects of professional responsibility and data privacy. During the feasibility analysis of the project (PROF-1 trial), the basic network model was defined by jointly identifying the individual professionals and their respective responsibilities. The investigators agreed that each of the professionals would be responsible for the accuracy of the

information provided, while the medication reconciliation process would be the exclusive responsibility of physicians. This decision meant that the data transmitted from a CP to IRST would only be valid after the clinician had downloaded, integrated, and confirmed the information. Regarding the CPs' task of collecting patient data, it was agreed that the pharmacist transmitting data directly to IRST medical records would not be able to view or modify the entered information.

A drug recognition form was created and validated in a pilot study carried out on breast cancer patients [18]. The form was divided into 5 sections (medications, critical foods, supplements, phytotherapeutic products, and homeopathic products). Information would be gathered for each compound (eg, active ingredient and trade name; pharmaceutical form and route of administration; dosage, posology, start or end of intake, reason for intake, and prescriber). Initially conceived as a paper document, the form was digitized in both the Dottorfarma IT platform and IRST medical records for the purposes of the trial.

A total of 200 patients were considered for the PROF-1 study. A formal sample size calculation for this prospective study was not performed due to the lack of preliminary data and the exploratory intent of the study. Eligible patients were required to meet the following criteria: being adults ≥18 years old, of either gender; having an Eastern Cooperative Oncology Group performance status of ≤1; undergoing anticancer treatment; having a clear understanding of the Italian language; and granting written informed consent.

When the patients came to the institute for anticancer treatment, health care professionals provided information about the trial and obtained written informed consent from those willing to take part. The patients were asked to choose one of the CPs accredited for the trial and to book an appointment for the pharmacological recognition step. The accredited pharmacist, after ensuring that the patient had provided written informed consent and completed a privacy form, interviewed the patient in a private consultation area, completing the online drug recognition form in the Dottorfarma IT platform (Figure 1), which has the same format as that of IRST electronic medical records. The pharmacist took into account the information provided by the patient or delegated caregiver and that retrieved from drug packages or other products used by the patient at home (and brought to the interview), paper referrals, or prescriptions from specialists. The recognition data were sent to IRST through the Dottorfarma IT platform and automatically saved in a specific medical history file in the electronic medical records (Figure 2). Upon confirmation of the next course of chemotherapy, the oncologist downloaded the pharmacological recognition form and performed the reconciliation process. For the purposes of the present trial, the processes of recognition and reconciliation were carried out only once for each enrolled patient. Following the completion of the drug recognition and reconciliation processes, the patients were asked to complete a satisfaction questionnaire.

Descriptive statistical analyses were performed on all case series (absolute and relative frequency for categorical variables and median and quantiles for continuous variables). All statistical analyses were carried out using SAS Software, version 9.4 (SAS Institute).

This study was reviewed and approved by the AUSL (Azienda unità sanitaria locale) Romagna/IRST Ethics Committee (approval number 1722, October 26, 2016) and was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and later versions. The participants provided written informed consent to take part in the study.

The data sets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Figure 1. Dottorfarma electronic scheme of the recognition form.

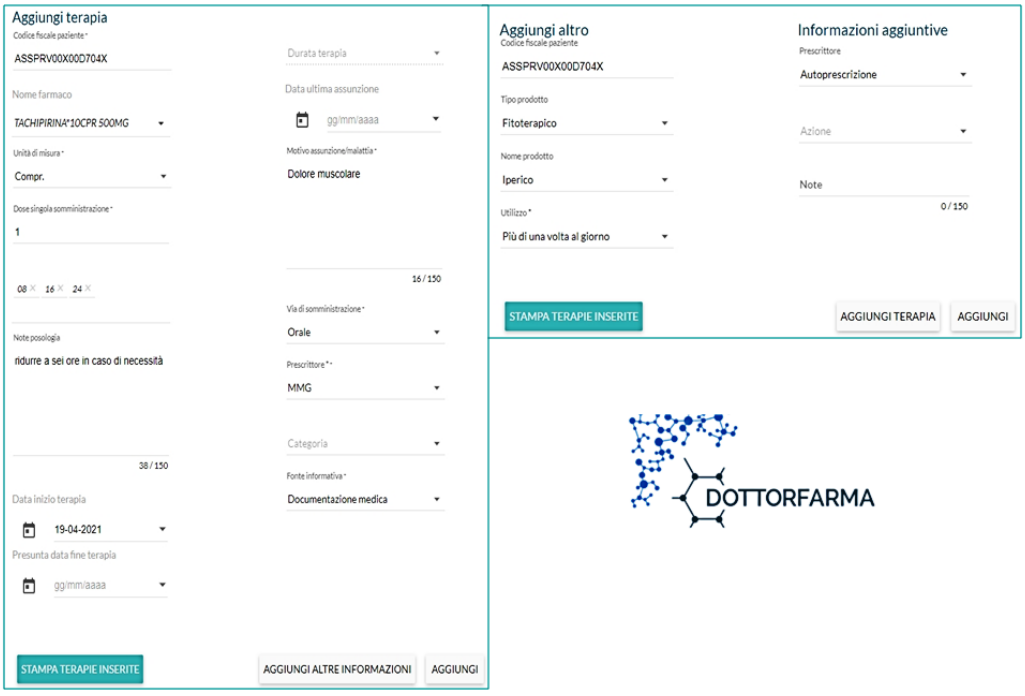
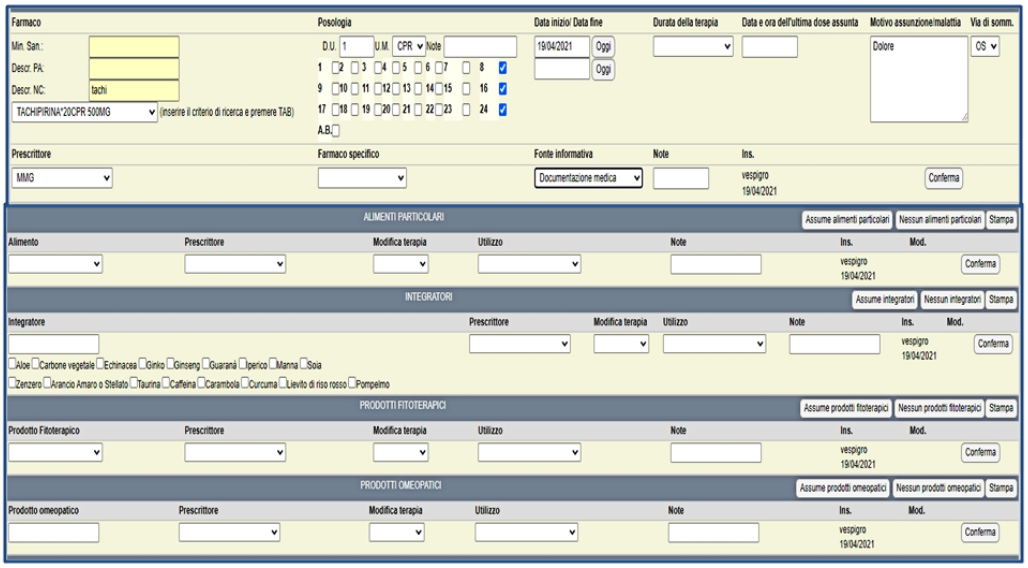


Figure 2. Drug history section of the IRST (Istituto Romagnolo per lo Studio dei Tumori) electronic medical records.



Results

A dialog between IRST and local CP associations was opened, leading to the creation of a specific agreement between the 2 parties. The first step involved the implementation of the Dottorfarma IT platform, initially with a one-way communication flow. The pharmacological recognition form was digitized on both the Dottorfarma IT platform and IRST electronic medical records, and the 2 systems were synchronized.

Pharmacy Recruitment and Training

Following an open invitation, out of 365, 120 (34%) of the public and private CPs of Romagna initially agreed to take part in the study. Pharmacy managers or a representative from each CP took part in a brief training course organized by IRST to inform the participants about the recognition and reconciliation process, to present the PROF-1 study, and to provide instructions on how to complete the pharmacological recognition form. A total of 108 CPs (84 private and 24 public CPs) finally agreed to participate in the trial (36 in the province of Forlì-Cesena, 26 in the province of Rimini, and 46 in the province of Ravenna). The enrolled patients chose 66 of these to complete the pharmacological recognition survey (Multimedia Appendix 1).

Patients

From April 2017 to April 2018, 200 patients were enrolled onto the trial. Among these, 137 (68.5%) underwent the pharmacological recognition interview as planned, while 63 (31.5%) were not evaluable, 39 (19.5%) withdrew consent, 11 (5.5%) were excluded for technical problems, and 13 (6.5%) spontaneously dropped out) (Multimedia Appendix 1). Patient characteristics are shown in Table 1.

The majority of the patients were female (83, 61%), with a median age of 63 years Most patients lived in the province of Forlì-Cesena, where our institute is based. The majority were undergoing chemotherapy (alone or in combination with targeted therapies), mainly administered intravenously. The list of anticancer drugs administered to the enrolled patients is reported in Table 2.

The analysis of the data entered by CPs revealed a total of 805 medications declared by the patients, with an average of 5.9 drugs per patient taken at home. As many as 12 or more medications were being taken by 6 patients over the age of 65 years. Drug distribution, identified in the recognition form according to the Anatomical Therapeutic Chemical classification, is reported in Table 3 (1st and 2nd levels).

Of the total 805 declared drugs, 29.4% (n=237) and 20.7% (n=167) were used to treat problems in the gastrointestinal and cardiovascular systems, respectively. The former drugs were mainly used to contrast chemotherapy side effects and cancer-related symptoms (eg, proton pump inhibitors and antacids, antiemetics, laxatives, and antidiarrheals), while the latter mainly included antihypertensive drugs and statins. Drugs for the central nervous system represented 13.2% (n=106) of all drugs and comprised mainly analgesics and antidepressants or anxiolytics. Hormones (55, 6.8%), drugs for musculoskeletal disorders (47, 5.8%), antithrombotic or antianemic drugs (43, 5.3%) and systemic antimicrobials (32, 4%) were also being taken. Although the recognition of anticancer drugs was not required (this information was already present in the medical records), some patients reported oral anticancer drugs in their list of medicines. Further interesting information was obtained from the recognition form; ie, 87.6% (n=705) of drugs were taken orally. Moreover, sources of information varied, coming from medical documents (145, 18%), drug lists (227, 28.2%), product packages (86, 10.7%), patient declarations (270, 33.5%), and caregiver declarations (73, 9.1%). The main prescribers were specialist physicians (542, 67.3%) and general practitioners (204, 25.3%), while 1.5% (n=12) of the patients were self-prescribers.

Table 1. Patient characteristics (n=137).

Variables	Values
Age (years), range (years)	67 (39-85)
Gender, n (%)	
Male	54 (39)
Female	83 (61)
Provence of residence, n (%)	
Forlì-Cesena	86 (63)
Ravenna	36 (26)
Rimini	8 (6)
Other	7 (5)
Site of disease , n (%)	
Gastrointestinal tract	36 (26)
Breast	31 (23)
Genitourinary tract	27 (20)
Hematologic malignancy	17 (12)
Lung	14 (10)
Anticancer treatment , n (%)	
Chemotherapy	55 (39)
Chemotherapy plus targeted therapy	28 (20)
Targeted therapy	21 (16)
Immunotherapy	13 (10)
Hormonal therapy	20 (15)
Administration route, n (%)	
Intravenous	74 (54)
Intravenous and oral	33 (24)
Oral	30 (22)

Table 2. Anticancer drugs (n=462).

Category and name of anticancer drugs	Values
Cytotoxic drugs, n (%)	
Paclitaxel	38 (8.2)
5-fluorouracil	36 (7.8)
Capecitabine	36 (7.8)
Cyclophosphamide	31 (6.7)
Gemcitabine	30 (6.4)
Oxaliplatin	29 (6.3)
Carboplatin	25 (5.4)
Cisplatin	23 (5.0)
Irinotecan	18 (3.9)
Docetaxel	17 (3.7)
Vinorelbine	14 (3.0)
Doxorubicin	13 (2.8)
Nab[a]-paclitaxel	11 (2.4)
Liposomal doxorubicin	10 (2.2)
Epirubicin	9 (1.9)
Bendamustine	8 (1.7)
Eribulin	7 (1.5)
Etoposide	7 (1.5)
Melfalan	7 (1.5)
Trifluridine+tipiracil	6 (1.3)
Pemetrexed	5 (1.1)
Total	380 (82.2)
Targeted therapy n, (%)	
Trastuzumab	14 (3.0)
Rituximab	11 (2.4)
Bevacizumab	11 (2.4)
Cetuximab	6 (1.3)
Regorafenib	7 (1.5)
Total	49 (10.6)
Immunotherapy n, (%)	
Nivolumab	13 (2.9)
Hormonal therapy n, (%)	
Fulvestrant	9 (1.9)
Abiraterone acetate	6 (1.3)
Enzalutamide	5 (1.1)
Total	20 (4.3)

[a]Nab: nanoparticle albumin-bound.

Table 3. Distribution of drugs (n=805) detected in the recognition (Anatomical Therapeutic Chemical, 1st and 2nd levels).

ATC[a] classification	Values
Gastrointestinal tract and metabolism, n (%)	
Proton pump inhibitors and antacids	82 (10.2)
Antiemetics and prokinetics	33 (4.1)
Vitamin A and D and associations	30 (3.7)
Insulin and hypoglycemic agents	24 (3.0)
Laxatives	18 (2.2)
Antidiarrheal and anti-inflammatory intestinal	17 (2.1)
Mineral supplements	14 (1.7)
Other	19 (2.3)
Total	237 (29.4)
Cardiovascular system, n (%)	
Antihypertensives (including diuretics)	128 (16.0)
Statins	34 (4.2)
Other	5 (0.6)
Total	167 (20.7)
Central nervous system, n (%)	
Analgesics (including opioids)	53 (6.6)
Antidepressants, anxiolytics, and sedatives	36 (4.5)
Other	17 (2.1)
Total	106 (13.2)
Hormones (excluding insulin and sex hormones), n (%)	
Systemic corticosteroids	42 (5.2)
Thyroid preparations	12 (1.5)
Other	1 (0.1)
Total	55 (6.8)
Antineoplastic agents and immunomodulators, n (%)	
Oral anticancer drugs	39 (4.8)
Intramuscular hormonal antagonists	11 (1.4)
Total	50 (6.2)
Musculoskeletal system, n (%)	
Antigout drugs	23 (2.8)
Anti-inflammatory drugs (including NSAIDs[b])	16 (2.0)
Bisphosphonates	6 (0.7)
Other	2 (0.3)
Total	47 (5.8)
Blood and hematopoietic organs, n (%)	
Antithrombotic drugs	32 (4.0)
Antianemic drugs	11 (1.3)
Total	43 (5.3)
Systemic antimicrobials, n (%)	
Antibacterials	22 (2.7)

ATC[a] classification	Values
Antivirals	9 (1.1)
Antifungals	1 (0.1)
Total	32 (4.0)
Genitourinary tract and sex hormones, n (%)	20 (2.5)
Respiratory system, n (%)	18 (2.2)
Skin, n (%)	7 (1.0)
Other, n (%)	23 (2.9)

[a]ATC: Anatomical Therapeutic Chemical.

[b]NSAIDs: nonsteroidal anti-inflammatory drugs.

In addition, of a total of 137 patients, 83 (60.5%) reported an intake of 201 nonconventional medications (supplements, phytotherapeutics, or homeopathic products), and 39 (28.5%) reported having taken foods considered critical for the potential interactions with medications (Table 4). A detailed analysis of the recognition form revealed that the most widely represented critical substances were coffee (23 patients, 16.8%), green tea (11 patients, 8%), aloe (6 patients, 4.4%), turmeric (5 patients, 3.6%), fermented red rice (5 patients, 3.6%), ginger (3 patients, 2.2%) and manna (3 patients, 2.2%).

An evaluation of drug-drug interaction was not carried out as it was not one of the aims of this paper. However, we investigated the potential for interaction between nonconventional products and cancer treatments (Table 4), identifying 2 possible interactions with aloe vera, 3 with manna, 1 with echinacea, 1 with ginseng, 2 with ginger, and 3 with red yeast rice. After analyzing the components present in the supplements taken by the patients but not present in our list of critical compounds, the following possible interactions emerged: 2 interactions with milk thistle, 2 with berberine, 1 with alpha lipoic acid, 1 with vitamin C, 1 with folic acid, and 1 with *Cordyceps sinensis*.

A total of 106 patients completed and returned the satisfaction questionnaire. Of these, 77 (72%) considered the reconciliation process as very important, 74 (70%) thought that the involvement of the pharmacist was very useful, and 87 (82%) reported no difficulty in going to the chosen pharmacy to complete the pharmacological recognition survey.

Table 4. Critical foods, nonconventional products, and potential interactions with cancer treatments.

Product	Value	Potential interaction
Critical foods, n (%)		
Coffee	23 (16.8)	—[a]
Green tea	11 (8)	—
Black tea	—	—
Bitter orange	—	—
Carom	—	—
Grapefruit	1 (0.7)	Regorafenib
Chili pepper	—	—
Pepper	—	—
Turmeric	5 (3.6)	Doxorubicin, cyclophosphamide
Phytotherapeutic, n (%)		
Aloe vera	6 (4.4)	Paclitaxel, docetaxel
Charcoal	—	—
Echinacea	1 (0.7)	Dexamethasone
Ginseng	1 (0.7)	Etoposide
Guarana	—	—
Hypericum	—	—
Red yeast rice	5 (3.6)	Cyclophosphamide, paclitaxel, etoposide
Manna	3 (2.2)	Capecitabine, enzalutamide, abiraterone
Soy	—	—
Ginger	3 (2.2)	5-fluorouracil, capecitabine
Other compounds[b], n (%)		
Milk thistle	2 (1.5)	Regorafenib, paclitaxel
Alpha-lipoic acid	1 (0.7)	Cisplatin
Vitamin C	1 (0.7)	Doxorubicin
Folic acid	1 (0.7)	Capecitabine
Berberine	2 (1.5)	Trastuzumab, vinorelbine
Cordyceps sinensis	1 (0.7)	Dexamethasone

[a]Not applicable.

[b]Not reported in the official list of critical compounds, but with possible critical interactions in the post hoc analysis.

Discussion

Principal Findings

The PROF-1 trial achieved its goal of creating a new model for medication recognition and reconciliation processes in oncology thanks to the close cooperation between our institute and CPs, the implementation of an integrated IT platform, and the active participation of cancer patients.

An interest in being actively involved in the setting up of a new hospital territory network was clearly demonstrated by both public and private CPs in their willingness to take part, at no cost, in the project. A total of 120 CPs participated in the training course, and 108 agreed to enrol patients, the latter process involving a commitment of around 30 minutes for each survey to integrate into routine pharmacy activities. To the best of our knowledge, this was the first trial addressing the problem of medication reconciliation to include private entities not directly involved in the care of cancer patients.

PROF-1 confirmed the high number of medications used by patients at home (an average of 5.9 drugs per patient), especially in the most advanced age groups. Even more striking was the evidence of an increasing number of patients who used nonconventional products or critical foods. These results underline the importance of pharmacological recognition and reconciliation processes, which, despite being mandatory ministerial measures [1,3], are often neglected by oncologists because of the pressures of daily clinical activity. An alliance with CPs could thus lead to a significant improvement in the situation [19-21]. Numerous attempts have been made to

improve the medication reconciliation process, including pharmacist-related interventions and IT platforms, with interesting results in terms of reduction in medication discrepancies and potential adverse events [22]. However, none of the proposed models included the presence of community pharmacists.

Our study highlighted some limitations in this new model, the first concerning patient empowerment. Although the patients were personally involved and expressed a high degree of satisfaction through the questionnaire, a certain number who signed the informed consent did not go to the chosen CP to complete the recognition step. This may have been due to a deterioration in their clinical conditions or to the side effects of chemotherapy, given that the study population was recruited among those undergoing antiblastic therapy, regardless of prognosis or type of treatment. This is an important aspect to bear in mind as it suggests that some patients may not be suitable for this type of project. A review of the communication channels between patient and CP, including the use of the telephone or email rather than direct access, could perhaps improve this issue. Another important limitation was that, given the exploratory nature of the trial, only one recognition-reconciliation process was planned for each patient, making it impossible to verify the advantage of a continuous exchange of information from a series of repeated processes to monitor changes in medications taken at home. A new trial (PROF-2) is ongoing to further implement and validate the model, this time incorporating the repetition of the recognition-reconciliation processes before each chemotherapy cycle. Another imitation concerns drug-drug interactions. As stated above, the use of a surprisingly high number of drugs, supplements, and nonconventional products were declared by patients, raising the question of possible interactions. However, we did not focus on the drug-drug interactions that emerged from the reconciliations performed by oncologists, as this was not an aim of the study. Conversely, our analysis of the potential interactions between cancer drugs and nonconventional products confirmed the importance of this issue. Finally, the difficult reproducibility of the proposed model must be emphasized as a possible limitation because of some basic requirements (ie, a high level of computerization of the centers taking part and the strict regulation of data processing and patient privacy). The strengths that made it possible to complete the project were the solidity of the computerized medical records present in our cancer institute, the presence of an IT platform that could be shared by all private and public CPs, and the approval of the project by the local ethics committee after in-depth teamwork to resolve the issues of professional responsibility, data ownership, and privacy management.

Based on the results from this pilot study, the alliance between the cancer center and the CPs of Romagna has led to the creation of a cancer network which, albeit initially established to meet a specific need, could help to systematically establish safety pathways for drugs used at home. This initiative could also contribute to increasing adherence to innovative drugs, essential not only for the success of treatment but also for the sustainability of the national health service of every nation. It could also improve the management of ancillary drugs for the prevention of toxicities, often underestimated or self-managed by patients [23]. The PROF-1 project was conceived and designed some years before the COVID-19 pandemic, which has forced the community to acknowledge the importance of the territory in the management of patients' needs. Within this context, our model proved highly functional not only in terms of the study's main aim (pharmacological reconciliation), but also in terms of the successful creation of a hospital community network whose impact may well exceed what was originally hypothesized.

Conclusions

The PROF-1 trial represents an important step forward in medication reconciliation in oncology. The alliance established between our cancer institute and local CPs to enhance medication reconciliation in transitions in care led to the creation of an innovative organizational model and the validation of a new integrated IT platform.

Acknowledgments

The authors wish to thank the late Professor Dino Amadori for his important contribution to the study design and for his support of this ambitious project from the very beginning. They also wish to thank all the patients who agreed to use the CPs (community pharmacy) involved in the study, thus contributing to the development of this experimental model; the CP pharmacists who took part in the project; Chiara Crispo and Andrea Tarozzi, Professor of Pharmacology at Bologna University, for their help with the analysis of drug interactions. Many authorities and professionals must be thanked for the support given to the project: the regional and provincial presidents of Federfarma in office at the start of the project and those currently in office (Daniele Raganato, Benedetta Camerani, Domenico Dal Re, Achille Gallina Toschi, Alberto Lattuneddu, Roberto Deluigi, and Pietro Gueltrini); Ernesto Toschi, the regional president of Assofarm, and Barbara Pesci and Bruna Baldassarri of the Municipal Pharmacy Company of Ravenna; Bruno Foresti of Federfarma and Giulia Fatano of Promofarma; Giovanni Petrosillo and Daniele D'Angelo, former directors of Promofarma, and Claudio Montalto, IT engineer formerly at Promofarma; Marco Cossolo, president of the National Federfarma, Venanzio Gizzi, president of Assofarm, and Annarosa Racca, the former president of the National Federfarma; Susanna Ciampalini of the Quality Office of the General Directorate of Health Planning, and Ester Sapigni from the Drug Policy Service of the Emilia Romagna Region.

Authors' Contributions

MVM, RV, and ON conceived the idea for and designed the study. MVM, AP, and RV collected patient data. AP, MVM, ON, PS, RV, CD, FF, SP, GT, PFG, GM, MA, and GLF were responsible for data interpretation. ON performed the statistical analyses. AP and MVM drafted the paper. All coauthors reviewed the paper, providing important feedback to improve the manuscript. All authors read and approved the present version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PROF-1 (Progetto di Rete in Oncologia con le Farmacie di comunità della Romagna) trial flow chart.
[DOCX File , 74 KB - jmir_v23i12e31321_app1.docx]

References

1. Raccomandazione n. 17. Riconciliazione della terapia farmacologica. Ministry of Health of Italy. 2014 Dec. URL: https://www.salute.gov.it/imgs/C_17_pubblicazioni_2354_allegato.pdf [accessed 2021-12-09]
2. Regione Emilia Romagna. Sicurezza nella terapia farmacologica "Processo di ricognizione e di riconciliazione farmacologica per una prescrizione corretta e sicura". Regione Emilia Romagna. 2015 Oct. URL: https://salute.regione.emilia-romagna.it/normativa-e-documentazione/rapporti/sicurezza-terapia-farmacologica/sicurezza_farmacologica_2-rev2015.pdf [accessed 2021-12-09]
3. Raccomandazione per la prevenzione degli errori in terapia con farmaci antineoplastici. Ministry of Health of Italy. 2012 Oct 14. URL: https://www.salute.gov.it/imgs/C_17_pubblicazioni_1861_allegato.pdf [accessed 2021-11-09]
4. No authors listed. Using Medication Reconciliation to Prevent Errors. The Joint Commission Journal on Quality and Patient Safety 2006 Apr;32(4):230-232. [doi: 10.1016/s1553-7250(06)32030-2] [Medline: 16649655]
5. Qato DM, Wilder J, Schumm LP, Gillet V, Alexander GC. Changes in Prescription and Over-the-Counter Medication and Dietary Supplement Use Among Older Adults in the United States, 2005 vs 2011. JAMA Intern Med 2016 Apr;176(4):473-482 [FREE Full text] [doi: 10.1001/jamainternmed.2015.8581] [Medline: 26998708]
6. The High 5s Project. Interim Report. World Health Organisation. 2013. URL: https://www.who.int/patientsafety/implementation/solutions/high5s/High5_InterimReport.pdf [accessed 2021-12-09]
7. Kwan JL, Lo L, Sampson M, Shojania KG. Medication reconciliation during transitions of care as a patient safety strategy: a systematic review. Ann Intern Med 2013 Mar 05;158(5 Pt 2):397-403 [FREE Full text] [doi: 10.7326/0003-4819-158-5-201303051-00006] [Medline: 23460096]
8. The High 5s Project. Standard Operating Protocol. Assuring Medication Accuracy at Transitions in Care. World Health Organisation. URL: https://www.who.int/patientsafety/implementation/solutions/high5s/h5s-sop.pdf [accessed 2021-12-09]
9. Daruwalla Z, Thakkar V, Aggarwal M, Kiasatdolatabadi A, Guergachi A, Keshavjee K. Patient Empowerment: The Role of Technology. Stud Health Technol Inform 2019;257:70-74. [Medline: 30741175]
10. Gobal Oncology Trends 2019. The IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/global-oncology-trends-2019 [accessed 2021-12-09]
11. Riechelmann RP, Tannock IF, Wang L, Saad ED, Taback NA, Krzyzanowska MK. Potential drug interactions and duplicate prescriptions among cancer patients. J Natl Cancer Inst 2007 Apr 18;99(8):592-600. [doi: 10.1093/jnci/djk130] [Medline: 17440160]
12. Alqenae FA, Steinke D, Keers RN. Prevalence and Nature of Medication Errors and Medication-Related Harm Following Discharge from Hospital to Community Settings: A Systematic Review. Drug Saf 2020 Jun;43(6):517-537 [FREE Full text] [doi: 10.1007/s40264-020-00918-3] [Medline: 32125666]
13. Bonaudo M, Martorana M, Dimonte V, D'Alfonso A, Fornero G, Politano G, et al. Medication discrepancies across multiple care transitions: A retrospective longitudinal cohort study in Italy. PLoS One 2018;13(1):e0191028 [FREE Full text] [doi: 10.1371/journal.pone.0191028] [Medline: 29329310]
14. Medication Without Harm. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255263/WHO-HIS-SDS-2017.6-eng.pdf;jsessionid=A93A39CC2B5A9DEEF2F43FFC0C2A70DB?sequence=1 [accessed 2021-12-09]
15. Monitoraggio della Spesa Farmaceutica Nazionale e Regionale Gennaio-Luglio 2020. Agenzia Italiana del Farmaco (AIFA). 2020 Dec 10. URL: https://www.aifa.gov.it/documents/20142/847405/Monitoraggio_Spesa_gennaio-luglio-2020.pdf/4170fed0-1430-dcfe-1e6c-53b7d6fa7592 [accessed 2022-12-09]
16. Rapporto Censis sul valore sociale dell'integratore alimentare. Centro Studi Investimenti Sociali (CENIS). 2019. URL: https://www.federsalus.it/wp-content/uploads/2019/06/Rapporto-Censis-definitivo-1.pdf [accessed 2021-12-09]
17. Axelsson R, Axelsson SB. Integration and collaboration in public health--a conceptual framework. Int J Health Plann Manage 2006;21(1):75-88. [doi: 10.1002/hpm.826] [Medline: 16604850]

18. Tenti E, Casadei Gardini A, Schirone A, Passardi A, Minguzzi M, Prati E, et al. Medication reconciliation in cancer day hospital. JCO 2015 May 20;33(15_suppl):e17748-e17748. [doi: 10.1200/jco.2015.33.15_suppl.e17748]
19. Mekonnen AB, McLachlan AJ, Brien JAE. Effectiveness of pharmacist-led medication reconciliation programmes on clinical outcomes at hospital transitions: a systematic review and meta-analysis. BMJ Open 2016 Feb 23;6(2):e010003 [FREE Full text] [doi: 10.1136/bmjopen-2015-010003] [Medline: 26908524]
20. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-based medication reconciliation practices: a systematic review. Arch Intern Med 2012 Jul 23;172(14):1057-1069 [FREE Full text] [doi: 10.1001/archinternmed.2012.2246] [Medline: 22733210]
21. Broadfield L, Shaheen P, Rogez M, Jamieson K, McCallum M. Guidelines for outpatient cancer care by community pharmacists. Can Pharm J (Ott) 2017;150(1):24-31 [FREE Full text] [doi: 10.1177/1715163516680009] [Medline: 28286590]
22. Mueller SK, Sponsler KC, Kripalani S, Schnipper JL. Hospital-based medication reconciliation practices: a systematic review. Arch Intern Med 2012 Jul 23;172(14):1057-1069 [FREE Full text] [doi: 10.1001/archinternmed.2012.2246] [Medline: 22733210]
23. Broadfield L, Shaheen P, Rogez M, Jamieson K, McCallum M. Guidelines for outpatient cancer care by community pharmacists. Can Pharm J (Ott) 2017;150(1):24-31 [FREE Full text] [doi: 10.1177/1715163516680009] [Medline: 28286590]

Abbreviations

AUSL: Azienda unità sanitaria locale
CP: community pharmacy
IRST: Istituto Romagnolo per lo Studio dei Tumori
IT: information technology
PROF: Progetto di Rete in Oncologia con le Farmacie di comunità della Romagna

Edited by R Kukafka; submitted 29.06.21; peer-reviewed by M Salimi, M Khokhar; comments to author 05.08.21; revised version received 21.09.21; accepted 29.10.21; published 20.12.21.

Please cite as:
Passardi A, Serra P, Donati C, Fiori F, Prati S, Vespignani R, Taglioni G, Farfaneti Ghetti P, Martinelli G, Nanni O, Altini M, Frassineti GL, Minguzzi MV
An Integrated Model to Improve Medication Reconciliation in Oncology: Prospective Interventional Study
J Med Internet Res 2021;23(12):e31321
URL: https://www.jmir.org/2021/12/e31321
doi:10.2196/31321
PMID:34932001

Original Paper

Implementation of Fingerprint Technology for Unique Patient Matching and Identification at an HIV Care and Treatment Facility in Western Kenya: Cross-sectional Study

Noah Kasiiti Jaafa[1*], BSc, MHI; Benard Mokaya[1*], BSc; Simon Muhindi Savai[1*], BSc; Ada Yeung[2], MHSA; Abraham Mosigisi Siika[3], MBChB, MMed; Martin Were[4*], MS, MD

[1]Institute of Biomedical Informatics, Moi University, Eldoret, Kenya

[2]Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Department of Medicine, School of Medicine, Moi University, Eldoret, Kenya

[4]Department of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[*]these authors contributed equally

Corresponding Author:
Noah Kasiiti Jaafa, BSc, MHI
Institute of Biomedical Informatics
Moi University
Nandi Road
Eldoret, PO Box 3900-30100
Kenya
Phone: 256 753764102
Email: kasiitinoah@gmail.com

Abstract

Background: Unique patient identification remains a challenge in many health care settings in low- and middle-income countries (LMICs). Without national-level unique identifiers for whole populations, countries rely on demographic-based approaches that have proven suboptimal. Affordable biometrics-based approaches, implemented with consideration of contextual ethical, legal, and social implications, have the potential to address this challenge and improve patient safety and reporting accuracy. However, limited studies exist to evaluate the actual performance of biometric approaches and perceptions of these systems in LMICs.

Objective: The aim of this study is to evaluate the performance and acceptability of fingerprint technology for unique patient matching and identification in the LMIC setting of Kenya.

Methods: In this cross-sectional study conducted at an HIV care and treatment facility in Western Kenya, an open source fingerprint application was integrated within an implementation of the Open Medical Record System, an open source electronic medical record system (EMRS) that is nationally endorsed and deployed for HIV care in Kenya and in more than 40 other countries; hence, it has potential to translate the findings across multiple countries. Participants aged >18 years were conveniently sampled and enrolled into the study. Participants' left thumbprints were captured and later used to retrieve and match records. The technology's performance was evaluated using standard measures: sensitivity, false acceptance rate, false rejection rate, and failure to enroll rate. The Wald test was used to compare the accuracy of the technology with the probabilistic patient-matching technique of the EMRS. Time to retrieval and matching of records were compared using the independent samples 2-tailed *t* test. A survey was administered to evaluate patient acceptance and satisfaction with use of the technology.

Results: In all, 300 participants were enrolled; their mean age was 36.3 (SD 12.2) years, and 58% (174/300) were women. The relevant values for the technology's performance were sensitivity 89.3%, false acceptance rate 0%, false rejection rate 11%, and failure to enroll rate 2.3%. The technology's mean record retrieval speed was 3.2 (SD 1.1) seconds versus 9.5 (SD 1.9) seconds with demographic-based record retrieval in the EMRS (P<.001). The survey results revealed that 96.3% (289/300) of the participants were comfortable with the technology and 90.3% (271/300) were willing to use it. Participants who had previously used fingerprint biometric systems for identification were estimated to have more than thrice increased odds of accepting the technology (odds ratio 3.57, 95% CI 1.0-11.92).

Conclusions: Fingerprint technology performed very well in identifying adult patients in an LMIC setting. Patients reported a high level of satisfaction and acceptance. Serious considerations need to be given to the use of fingerprint technology for patient

identification in LMICs, but this has to be done with strong consideration of ethical, legal, and social implications as well as security issues.

(J Med Internet Res 2021;23(12):e28958) doi:10.2196/28958

KEYWORDS

biometrics; patient matching; fingerprints; unique patient identification; electronic medical record systems; low- and middle-income countries (LMICs)

Introduction

Background

Unique patient matching and identification across health services is an operational challenge in many health care settings [1-4]. Commonly used identifiers such as patient names, ID numbers, national ID numbers, and dates of birth are often inadequate at guaranteeing unique patient identification, and sometimes these data may not be available [5-7]. Failure to correctly match patients largely contributes to inefficiencies in care delivery, resulting in medical errors that can affect patient safety [8]. In the United States alone, an estimated average of 400,000 deaths occur annually because of medical errors [9]. The problem is even more significant in low- and middle-income countries (LMICs), with many of these countries yet to implement accurate and comprehensive person identification procedures that ensure unambiguous identification of all citizens and residents [10].

Patient identity management has relied heavily on either deterministic or probabilistic algorithms, as well as other statistical matching procedures [11,12]. Deterministic matching algorithms use exact match or comparisons between 2 fields [13]. These algorithms compare identifiers such as national IDs for each record to determine a match. Probabilistic matching, which is by far the most widely implemented technique for record matching [14], does not necessarily depend on unique identifiers [15]. This technique assigns weights to records to calculate the probability of linkage among them; it then uses the observed frequency of agreement and disagreement patterns among the records. The total linkage weight is then compared with a threshold above which pairs are considered a match [16-18]. The challenge with deterministic algorithms is their lack of scalability and the requirement for expensive customization and business rule revisions as databases grow in size [19]. Furthermore, not all probabilistic algorithms applied to the same set of circumstances yield results with the same degree of accuracy because they are based on field-specific weights rather than value-specific weights [20]. Application of statistical matching approaches is further limited where patients may have isolated sets of their information distributed across multiple disparate systems. In addition, people may not consistently use their official names, with some preferring to use abbreviations or alternative expressions. Name misspellings and name-order transpositions are also encountered [21,22].

In our own institutional experience, the evaluation of various 4-string manipulation strategies to improve the performance of probabilistic models of patient matching based on Kenyan names revealed a suboptimal specificity and positive predictive value of <50% when matching patient names from 2 independent clinical databases [23]. This was indeed a low level of performance accuracy, especially in a health care setting. It is therefore imperative that better patient-matching approaches are determined, especially as settings start to exchange, aggregate, and centralize clinical data from multiple sources. Beyond matching performance, any approach to be considered has to be relatively affordable, feasible, and acceptable in LMIC care settings.

Use of Fingerprinting

Biometric approaches offer a potential solution to the challenge by trying to tackle some of issues identified in the traditional patient-matching algorithms [24]. The use of fingerprinting is already quite common in many LMICs, often for national-level voter registration and immigration services, among others. Fingerprinting and other biometric technologies are not without challenges. Special consideration needs to be given to the ethical, cultural, social, and legal implications of these solutions [25]. For example, some people may have concerns about the safety, encryption, and secondary use of their biometric data [26]. Studies have cited concerns by users about the hygiene of touching these devices [27-29]. Given the potential benefits of these technologies for improving patient matching and the poor performance of existing statistical patient-matching models, rigorous evaluations need to be conducted that weigh fingerprinting benefits (improving matching) against concerns by key stakeholders.

The basic principle of biometric authentication is that everyone is unique and can be identified by their intrinsic physical or behavioral traits [30]. Biometric technologies such as fingerprinting offer a potentially promising solution given that fingerprinting scanners are ubiquitous, the technology is relatively easy to acquire and use [31], the scanners require minimal database memory requirements, and there have been several demonstrated instances of their large-scale implementation and dominance in other sectors [32,33]. A fingerprint consists of a pattern of ridges and valleys in the surface of the fingertips, and its formation is related to the earlier fetal months [34]. It is the ridges and valleys that are extracted as minutiae that are converted into a template for storage in the database. The template is always retrieved and compared with a fresh scan (new template) during matching and identifications processes [35]. In this study, we set out to (1) evaluate the performance of using fingerprint biometric technology for patient matching within an LMIC setting and (2) evaluate acceptability and patient perception of using this system for patient identification and matching.

Methods

Study Setting and Participants

This study was conducted at care facilities served by the Academic Model Providing Access to Healthcare (AMPATH) in Western Kenya. The program offers HIV care and treatment services, with more than 100,000 patients who have tested positive for HIV currently under care. The program operates in partnership with Kenya's Ministry of Health and with support from the United States Agency for International Development. This study was conducted at AMPATH clinical facilities located in Eldoret, Kenya. The study population included consenting patients aged ≥18 years visiting the AMPATH clinics for care. Patients aged <18 years were excluded because of the intricacies involved in obtaining their consent and in filling surveys. AMPATH uses an adaptation of the Open Medical Record System, an open source electronic medical record system (EMRS) deployed widely in LMICs [36,37]. Patient records in the system contain demographic information that includes several patient attributes, for example, name, age or birthdate, national ID number if available, and telephone number. The Fellegi–Sunter algorithm is used within the EMRS for patient matching. The Fellegi–Sunter algorithm is a weighted probabilistic-based record linkage approach that relies on matching weights to compute a maximum likelihood estimate that determines whether a record pair is a match or nonmatch [38].

Study Design

The mUzima Fingerprint Module version 1.0 was developed to work as an integrated module within the EMRS in a client server–model architecture. The module was developed using Java Web Start (Oracle Corporation), which uses the Java Network Launching Protocol to download and run the fingerprint application locally on the client machine. This particular technology provides an easy 1-click activation of remotely hosted applications, guaranteeing application efficiency and elimination of complicated installation or upgrade procedures [39].

Fingerprint Technology–Matching Algorithm

The fingerprint-matching technology was based on the Bozorth3 fingerprint-matching algorithm (National Institute of Standards and Technology) [40]. Bozorth3, a minutiae-based algorithm, does both one-to-one and one-to-many matching operations. The recommended Bozorth3 threshold is the integer 40. However, during the technology testing phase, test runs at a threshold of 40 often produced false positives, and after numerous tests, a threshold of 70 proved to be a better measure to work with for this study. For every patient, 3 instances of the same left thumbprint were captured, converted into templates, and stored in the database. A patient was identified positively if any of the 3 instances returned a matching value of ≥70, the set threshold; otherwise, it returned a nonmatch.

Fingerprint Technology User Interface

The fingerprint module user interface provided 2 options for searching a patient either by their demographic information (name) or by fingerprint. A name search using the probabilistic patient-matching algorithm of the current EMRS returned all database instances with the same or similar name, with matches refined as more information was entered (Figure 1). However, when the same person was searched using their fingerprint as the identifier, a distinct and unique result was returned (Figure 2). The workflow for the fingerprint technology module is presented in Multimedia Appendix 1.

Figure 1. Screenshot showing the output of a patient search ("Mark") using a probabilistic-based algorithm.

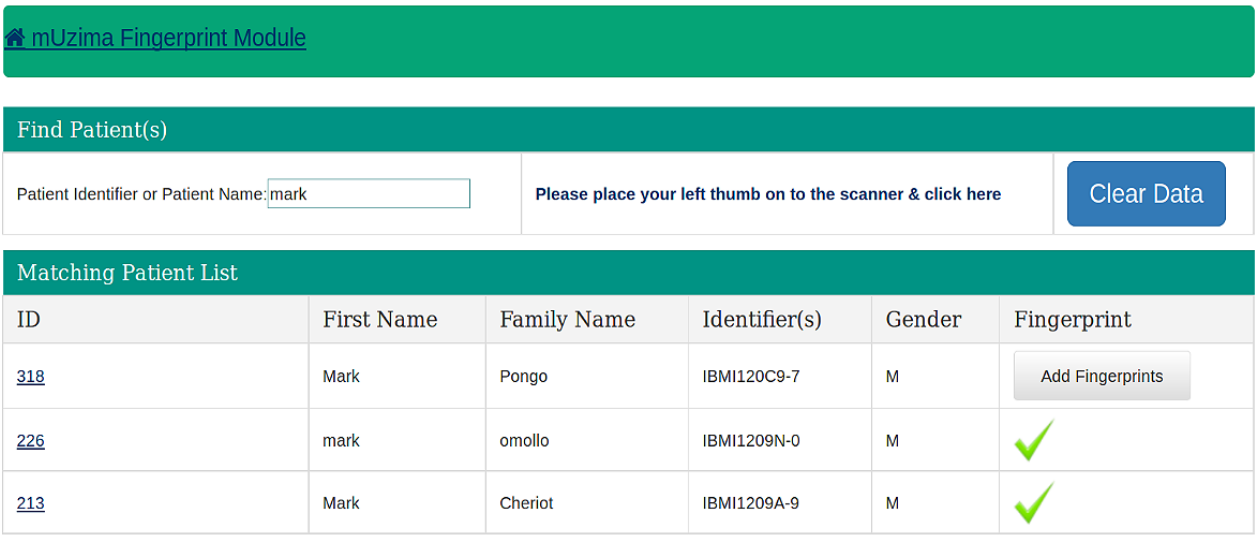
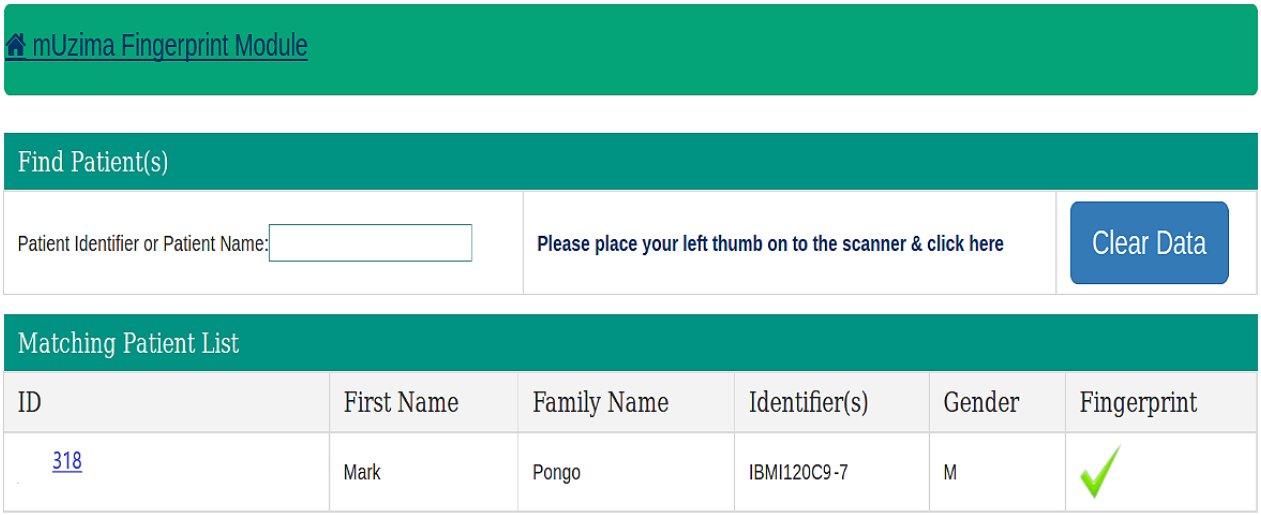


Figure 2. Screenshot showing the output of a patient search ("Mark") using fingerprint technology.



Matching Accuracy and Time to Retrieve Records (Fingerprint Technology vs Probabilistic Matching)

The current probabilistic patient-matching method used within the AMPATH EMRS was compared with the integrated fingerprint-matching and identification module. A convenience sample of 300 patients was recruited and enrolled for the study. When the patients arrived at the clinic for the visit, they received informed consent from the study researcher (NJK), and if they consented, 3 instances of their left thumbprint were captured through the EMRS interface. The fingerprint module converted the captured prints into a template and stored them in the database for later use, to either match or retrieve records. Record retrieval was performed at different points of care within the facility, as would have been required of a provider retrieving a patient record before offering care.

The performance of the fingerprint technology was evaluated using standard measures: participants correctly identified, sensitivity, false acceptance rate (FAR), false rejection rate (FRR), and failure to enroll rate (FER). The Wald test for proportions was used to compare the accuracy of the fingerprint technology with that of the Fellegi–Sunter probabilistic patient-matching technique used within the EMRS. Time to matching and retrieval of records (measured in seconds using a timer) were compared when records were retrieved through matching of data entered by clinical personnel against those retrieved through fingerprint technology–based patient matching. A survey was also administered to evaluate participants' acceptance and satisfaction of the fingerprint technology (Multimedia Appendix 2). The survey evaluated participants' understanding, comfort, perception, and willingness to accept the use of fingerprint technology for patient identification.

Fingerprint Technology Performance Measures (FAR, FRR, and FER)

Data entry and analyses were performed using STATA software (version 13.1 SE; StataCorp LLC). Age was summarized using mean and the corresponding SD. Categorical variables, for example, gender, level of education, and occupation, were summarized using frequencies and the corresponding percentages. Calculations were also performed on the standard biometric measures (Table 1).

Table 1. Calculated fingerprint technology performance measures.

Measure	Definition	Formula
Sensitivity	The ability of a system to correctly identify and match patients enrolled in the database	True positives/(true positives + false negatives) × 100
Specificity	The ability of the system to correctly reject patients not enrolled in the database	True negatives/(true negatives + false positives) × 100
False acceptance rate	The probability that a user who should be rejected is accepted by the system	False acceptance/total number of attempts × 100
False rejection rate	The probability that a user who should be accepted is rejected by the system	False rejection/total number of attempts × 100
Failure to enroll rate	The rate at which attempts to create a template from a scanned image are unsuccessful	False attempts/total number of attempts × 100

Acceptability of the Fingerprint-Matching System

Acceptability of the fingerprint biometric system was derived using 3 stand-alone variables: willingness to use the fingerprint system in the future, threats to patient privacy as a result of using the biometric system, and comfort while using the system during registration. The association between categorical variables and acceptability was assessed using the Pearson chi-square test. The logistic regression model was used to assess the predictors of acceptability of the fingerprint biometric system. Odds ratios and the corresponding 95% CIs were reported. The Wald test for proportions was used to compare the level of accuracy of the fingerprint technology with the null value (level of accuracy reported of the probabilistic patient-matching approach).

This study was approved by the institutional review and ethics committee at Moi University in Eldoret, Kenya (MU/MTRH-IREC approval number FAN: IREC 1832).

Results

Overview

A total of 307 participants who were approached consented to the study, of whom 300 (97.7%) were registered using the fingerprint technology for enrollment into the study; the other 7 (2.3%) could not be enrolled because of difficulties in acquiring their prints, thus leading to an FER of 2.3%.

Demographic details of those enrolled are outlined in Table 2. The participants' mean age was 36.3 (SD 12.2) years. Of the 300 participants, 174 (58%) were women and 215 (71.7%) had secondary or higher level of education (Table 2).

Table 2. Sociodemographic characteristics of participants (N=300).

Variable	Values
Age (years), mean (SD)	36.1 (12.3)
Gender, n (%)	
Male	126 (42)
Female	174 (58)
Education level, n (%)	
Informal	10 (3.3)
Primary	75 (25)
Secondary	78 (26)
Tertiary (college or university graduate)	137 (45.7)
Occupation, n (%)	
Student	66 (22)
Working or retired	125 (41.7)
Informal employment	109 (36.3)

Fingerprint Technology Performance

Considering the 300 appropriately scanned fingerprints, the system had a sensitivity of 89.3% (268/300) and a specificity of zero. However, the 32 unidentified fingerprints during the patient-matching exercise were included as false negatives (32/300, 10.7%; Table 3).

Table 3. False acceptance rate (FAR), false rejection rate (FRR), and failure to enroll rate (FER) for the fingerprint technology.

Metric	Calculation	Result (%)
FER	7/307 × 100	2.3
FAR	0/300 × 100	0
FRR	32/300 × 100	10.7
Sensitivity	268/(266 + 32) × 100	89.3
Specificity	0/(0 + 0) × 100	0

FAR, FRR, and FER

The fingerprint technology had an FAR of 0%, signifying that the technology was able to perfectly determine that an individual was not yet enrolled into the system, returning such instances as nonmatches. However, of the 300 enrolled individuals, 32 (10.7%) were falsely considered nonmatches (FRR), signifying that some individuals who were in the system could sometimes not be matched (Table 3). The reasons for the slightly high FRR were as follows: (1) weak finger pressure on the scanner (12/32, 37%), (2) faded prints (6/32, 19%), (3) scars on the thumb (6/32, 19%), (4) scanner response failure error (5/32, 16%), and (5)

low-quality images (3/32, 9%). Of these reasons, (4) and (5) were system-generated notifications, whereas the rest were manually observed and tallied by the research team during the exercise.

Matching Accuracy and Time to Retrieve Records (Fingerprint Technology vs Probabilistic Matching)

The accuracy of the fingerprint technology compared with that of the already established accuracy levels of the EMRS that ranged from a minimum of 50% to a maximum of 70% showed a value of $P<.001$ at both minimum and maximum levels, with a power of >99%. It took an average of 3.2 (SD 1.1, range 1.0-6.8) seconds to retrieve patient records using the fingerprint technology compared with 9.5 (SD 1.9, range 6.0-13.0) seconds ($P<.001$) using the probabilistic algorithm of the AMPATH Medical Record System.

Participants' Perception of Biometric Systems

Of the 300 participants, only 107 (35.7%) had heard of a biometric system before; 34 (11.3%) had used a fingerprint identification system before and 11 (3.7%) a facial recognition system, whereas none had used either an eye or iris identification system or a voice-based identification system. Of the 300 participants, 255 (85%) either strongly agreed (65.7%, 197/300) or agreed (19.3%, 58/300) that a biometric system could improve identification, 30 (10%) disagreed, and 15 (5%) responded "Don't know" to this question. Of the 300 participants, 289 (96.3%) were comfortable or very comfortable with using the fingerprint system, 290 (96.7%) expressed satisfaction with the system, and 271 (90.3%) were willing to use fingerprint technology for patient matching and identification in the future (Table 4). Most (241/300, 80.3%) of the participants did not think that technology threatened their privacy.

Table 4. Participants' perception of the piloted fingerprint biometric system (N=300).

Variable	Values, n (%)
Level of comfort during registration using fingerprint system	
Very comfortable	228 (76)
Comfortable	61 (20.3)
A little comfortable	5 (1.7)
Not comfortable	3 (1)
Would rather not say	3 (1)
General perception of the system	
Satisfied	290 (96.7)
Dissatisfied	10 (3.3)
Willing to use fingerprint system in future	
Yes	271 (90.3)
No	8 (2.7)
Don't know	21 (7)
Technology threatens own privacy	
Yes	45 (15)
No	241 (80.3)
Don't know	14 (4.7)

Acceptability of Fingerprint Matching

Acceptability of the fingerprint technology was derived from 3 variables: participants' comfort: 96.3% (289/300), willingness to use the fingerprint technology in the future: 90.3% (271/300), and participants' concerns regarding their privacy: 15% (45/300). The variables used to denote acceptability were further combined, that is, acceptability=yes if a participant was willing to use the fingerprint biometric system in the future, was comfortable or very comfortable using the fingerprint biometric system during registration, and did not feel that their privacy was threatened. The derived composite revealed that up to 77% (231/300) of the participants demonstrated acceptability of the fingerprint biometric system.

The Pearson chi-square test demonstrated no association between sociodemographic characteristics and acceptability of the fingerprint technology. However, the findings revealed that the respondents who agreed that the fingerprint technology offered an improved solution for patient matching were associated with the acceptability of the fingerprint technology for future use, 87.5% (202/300) versus 76.8% (53/300; $P=.03$). Similarly, the respondents who were satisfied with the fingerprint technology for identification were also associated with the acceptability of the fingerprint system for future use, 99.1% (229/300) versus 88.4% (61/300; $P=.001$; Table 5).

After adjusting for the opinion on whether the fingerprint biometric system offered an improved solution, opinion on whether the current patient identifiers (eg, IDs and patient numbers) offered an appropriate solution for identification,

previous knowledge of biometric systems, and occupation, participants who had previously used fingerprint biometric systems for identification were estimated to have more than thrice the increased odds of accepting the fingerprint biometric system (odds ratio 3.57, 95% CI 1.0-11.92; Table 6).

Table 5. Bivariate associations of participants' perception of biometrics and acceptability of the fingerprint technology system (N=300).

Variables	Acceptable in the future, n (%)		*P* value
	No (n=69)	Yes (n=231)	
Have ever heard of biometric systems before			
No	42 (60.9)	151 (65.4)	—[a]
Yes	27 (39.1)	80 (34.6)	.49
Have ever used fingerprint biometric system before			
No	65 (94.2)	203 (87.9)	—
Yes	4 (5.8)	28 (12.1)	.14
Opinion on whether the current patient identifiers (eg, IDs and patient numbers) offer an appropriate solution for identification			
Agree or strongly agree	30 (43.5)	79 (34.2)	.16
Disagree or strongly disagree or don't know	39 (56.5)	152 (65.8)	—
Opinion on whether biometric systems offer improved solution			
Agree or strongly agree	53 (76.8)	202 (87.5)	—
Disagree or strongly disagree or don't know	16 (23.2)	29 (12.6)	.03[b]
General perception of the system			
Dissatisfied	8 (11.6)	2 (0.9)	—
Satisfied	61 (88.4)	229 (99.1)	.001[b]

[a]Not available.

[b]Pearson chi-square test.

Table 6. Determinants of acceptability of the fingerprint biometric system (N=300).

Variables	UOR[a] (95% CI)	AOR[b] (95% CI)
Biometric systems offer improved solution		
Disagree or strongly disagree or don't know	Reference	Reference
Agree or strongly agree	2.10 (1.06-4.16)	2.30 (1.12-4.69)
Current patient identifiers offer an appropriate solution for identification		
Disagree or strongly disagree or don't know	Reference	Reference
Agree or strongly agree	0.68 (0.39-1.17)	0.61 (0.34-1.09)
Have ever heard of biometric systems before	0.82 (0.47-1.43)	0.51 (0.27-0.96)
Have ever used fingerprint biometric system for identification	2.24 (0.76-6.63)	3.57 (1.07-11.92)[c]
Occupation		
Working or retired vs student	1.30 (0.63-2.71)	1.56 (0.73-3.33)
Informal employment vs student	0.74 (0.36-1.51)	0.72 (0.34-1.53)

[a]UOR: unadjusted odds ratio.

[b]AOR: adjusted odds ratio.

[c]Pearson chi-square test.

Discussion

Principal Findings

The matching speed of the fingerprint technology was significantly lower and better than that of the demographic-based retrieval matching technology, that is, 3.2 (SD 1.1) seconds versus 9.5 (SD 1.9) seconds (P<.001). The system operated in a mode where no imposter was falsely considered a match in 300 attempts. However, 10.7% (32/300) of the genuine attempts were falsely considered as nonmatches (FAR 0/300 = 0.0% at an FRR 32/300 = 10.7%). The participants' acceptability of the technology was demonstrated with most (289/300, 96.3%) of them indicating that they were comfortable with the technology and 90.3% (271/300) being willing to use the technology in the future. Although concerns about performance and acceptability of biometric technologies, especially fingerprint technology, are often discussed, few studies have comprehensively evaluated this technology in LMICs, despite fingerprint technology being widely available and affordable in these settings. Existing studies have largely described development of fingerprint technology systems without reporting on the performance using standard measures [41,42]. The few that have evaluated performance have not looked at the perceptions and acceptability of these systems [43]. To our knowledge, this is the first study to comprehensively evaluate performance, acceptability, and attitudes toward the use of fingerprint technology in an LMIC setting. The accuracy of the fingerprint technology–matching algorithm was significantly higher than that of the existing Fellegi–Sunter probabilistic algorithm. This is in line with the previous demonstration of the accuracy of fingerprint systems through systematic reviews, for example, a study of this technology in Nigeria recorded a success rate of 94% [43]. Not only did we demonstrate that fingerprint technology performed well, but we also demonstrated that it was more efficient than the traditional record-searching mechanisms at retrieving patient records from the EMRS. Data from this study add to the body of evidence of the potential of fingerprint technology to improve patient identification and matching in LMICs, which often lack unique identifiers and which experience other challenges to patient identification.

Despite performing better than traditional patient-matching approaches, the technology still failed to enroll a small number of individuals (FER 2.3%, 7/300) and to match 10.7% (32/300) of the registered individuals (FRR 10.7%, 32/300). A key reason for the failure to identify 10.7% (32/300) of the registered individuals was related to issues with the thumbprint of the individual patient (worn out thumbprint), signifying that performance of these systems could be improved by considering alternative or multiple fingers when one did not suffice. The other reasons were related to operator and technology issues, also suggesting that improving the scanner and software could improve system performance. To reduce costs, we tested the technology using only 1 scanner type (DigitalPersona U.are.U 4500; DigitalPersona, Inc) with open source software, but we recognize that there are other scanners that could perform better. There also needs to be better training of operators and improved level of machine-generated notifications to improve performance. In the COVID-19 pandemic era, it is important to not only understand how to operate these systems beyond simply recording fingerprints, but to also learn how to ensure that scanners are cleaned well and social distancing is maintained between the operator and the patient whose fingerprint is being captured. The aforementioned findings also suggest that fingerprint technology should not necessarily be seen as a replacement for traditional deterministic and probabilistic methods; rather, it should be looked upon as an additional and complementary enhancement.

This study indicated that fingerprint technology would be acceptable to patients in these settings. A surprisingly high number of participants (271/300, 90.3%) were willing to use the technology and expressed satisfaction with its use (290/300, 96.7%). This aligns with the findings from another study on adoption of fingerprint scanning during contact investigation for tuberculosis in Kampala, Uganda, where fingerprint technology was found to be feasible and acceptable [44]. A notable finding in our study was that participants' sociodemographic characteristics and their perceptions were generally not associated with acceptability of the fingerprint technology. Previous users of the fingerprint technology biometric system were associated with more than thrice the increased odds of accepting the fingerprint biometric solution, signifying that if implemented well, fingerprint technology is likely to achieve even more acceptance with patients.

Only 15% (45/300) of the participants felt that the technology threated their privacy. Previous studies have demonstrated similar lower concern regarding privacy in LMICs. As an example, in a study evaluating authentication options for mobile health apps in younger and older adults, only approximately 10% of the participants were concerned about the storage of their face scans, whereas only 4% were concerned about the privacy requirements [45]. Users' concerns regarding privacy tend to be paradoxical because they sometimes vary depending on the technology and user knowledge of the subject [46]. Despite our participant responses, privacy remains a key concern with fingerprint technology. In LMICs, there is a particular need for patient education and sensitization on the privacy and security of these technologies. Furthermore, there should be strong frameworks, guidelines, and standard operating procedures that cover ownership, control, and use of, as well as access to, patients' biometric data [47]. In fact, in settings such as those of LMICs, more protections are needed, given that patients might be vulnerable simply because they may have less understanding of the technology. As previously described, ethical guidelines to consider while implementing biometrics should include, but not be limited to, the following: (1) secondary principles and rights (right of access to information and data protection principles), (2) surveillance society issues (choice, power, empowerment, transparency and accountability, consent, and communication with the user), and (3) health and hygiene concerns regarding biometrics (physical contact between people providing the biometric data and the official enrolling the individuals) [48].

This study included several limitations that deserve a mention. The technology was tested on only 1 scanner type (DigitalPersona U.are.U 4500), and only the thumbprint was used. Although the system performed very well, there is room

for improvement to incorporate the capture of more fingers and to leverage the features of other scanners. In addition, infants were not catered for in this study, given the known challenges with fingerprint capture in this population [49,50]. Beyond attitudes and perceptions of patients, it would also be important to capture the perceptions of other stakeholders regarding fingerprint technologies, including providers, managers, and health ministry leaders. The next steps include further refinement of this system and working with key stakeholders to determine its role in supporting national health information systems in Kenya and other countries.

Conclusions

Fingerprint biometric technology offers an improved patient identification mechanism compared with traditional deterministic and probabilistic techniques. Accuracy can be greatly improved when more than one finger is captured during enrollment because this captures more minutiae. The technology is also acceptable to patients in resource-limited settings. However, large-scale deployment should take ethical, legal, and social implications as well as security issues into consideration.

Acknowledgments

This work was supported in part by the Norwegian Programme for Capacity Development in Higher Education and Research for Development program (Norwegian Agency for Development Cooperation: Project #QZA-0484, Health Informatics Training and Research in East Africa for Improved Health Care project). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Norwegian Agency for Development Cooperation. The authors would like to appreciate the patients who spared their time to share their experiences and opinions. Thanks especially to Academic Model Providing Access to Healthcare Kenya for granting the authors permission to use the facility as the study site.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fingerprint capture workflow.
[PNG File , 105 KB - jmir_v23i12e28958_app1.png]

Multimedia Appendix 2
Questionnaire to evaluate patient perception of fingerprinting patient identification system.
[DOCX File , 55 KB - jmir_v23i12e28958_app2.docx]

References

1. Morris G, Farnum G, Afzal S, Robinson C, Greene J, Coughlin C. Patient identification and matching final report. Audacious Inquiry, LLC - Prepared for the Office of the National Coordinator for Health Information Technology. 2014. URL: https://www.healthit.gov/sites/default/files/patient_identification_matching_final_report.pdf [accessed 2021-11-24]
2. Cidon D. Why patient identification remains a daunting challenge for healthcare. Healthcare Innovation. 2018. URL: https://tinyurl.com/2tn9373d [accessed 2021-11-24]
3. Gliklich RE, Dreyer NA, Leavy MB. Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014:1-356.
4. Gliklich R, Dreyer N, Leavy M. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Vol 2. Maryland: Agency for Healthcare Research and Quality; 2010:38-39.
5. Riplinger L, Piera-Jiménez J, Dooling JP. Patient identification techniques - approaches, implications, and findings. Yearb Med Inform 2020 Aug;29(1):81-86 [FREE Full text] [doi: 10.1055/s-0040-1701984] [Medline: 32823300]
6. Fernandes L, O'Connor M. Accurate patient identification - a global challenge. Perspect Heal Inf Manag. 2015. URL: https://perspectives.ahima.org/accurate-patient-identification-a-global-challenge/ [accessed 2021-11-24]
7. Cylus J, Papanicolas I, Smith PC. Health System Efficiency: How to Make Measurement Matter for Policy and Management. Geneva: World Health Organization; 2016:1-242.
8. Landi H. For hospitals, patient misidentification impacts patient safety, productivity and the bottom line. Healthcare Innovation. 2016. URL: https://tinyurl.com/5f5ua36u [accessed 2021-11-24]
9. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. Br Med J 2016 May 03;353:i2139. [doi: 10.1136/bmj.i2139] [Medline: 27143499]
10. Verbeke F, Karara G, Van Bastelaere S, Nyssen M. Patient identification and hospital information management systems in Sub-Saharan Africa: a prospective study in Rwanda and Burundi. Rwanda Med J 2012;69(4):1-6 [FREE Full text]
11. Hagger-Johnson G, Harron K, Goldstein H, Aldridge R, Gilbert R. Probabilistic linkage to enhance deterministic algorithms and reduce data linkage errors in hospital administrative data. J Innov Health Inform 2017 Jun 30;24(2):891 [FREE Full text] [doi: 10.14236/jhi.v24i2.891] [Medline: 28749318]

12. Just BH, Marc D, Munns M, Sandefer R. Why patient matching is a challenge: research on master patient index (MPI) data discrepancies in key identifying fields. Perspect Health Inf Manag 2016;13:1e [FREE Full text] [Medline: 27134610]
13. Schumacher S. Probabilistic versus deterministic data matching: making an accurate decision. Semantic Scholar. 2010. URL: https://www.semanticscholar.org/paper/Probabilistic-Versus-Deterministic-Data-Matching-%3A-Schumacher/743a6f53aa08eb0b89b3c4108c2ab36765658998 [accessed 2021-11-24]
14. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. Int J Epidemiol 2016 Jun 20;45(3):954-964 [FREE Full text] [doi: 10.1093/ije/dyv322] [Medline: 26686842]
15. Verykios VS, Elmagarmid AK, Houstis EN. Automating the approximate record-matching process. Inform Sci 2000 Jul;126(1-4):83-98. [doi: 10.1016/s0020-0255(00)00013-x]
16. Elezaj O, Tuxhari G. Record linkage using probabilistic methods and data mining techniques. Mediterr J Soc Sci 2017;8(3):203-207. [doi: 10.5901/mjss.2017.v8n3p203]
17. Dusetzina B, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR. Linking Data for Health Services Research: A Framework and Instructional Guide. Maryland: Agency for Healthcare Research and Quality and U.S. Department of Health and Human Services; 2014:1-86.
18. Machado CJ, Hill K. Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. Cad Saude Publica 2004 Aug;20(4):915-925 [FREE Full text] [doi: 10.1590/s0102-311x2004000400005] [Medline: 15300283]
19. Pita R, Menezes L, Barreto ME. Applying term frequency-based indexing to improve scalability and accuracy of probabilistic data linkage. LADaS 2018 - Latin America Data Science Workshop. 2018. URL: http://ceur-ws.org/Vol-2170/paper9.pdf [accessed 2024-11-21]
20. Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. J Am Med Inform Assoc 2009 Sep 01;16(5):738-745. [doi: 10.1197/jamia.m3186]
21. Joffe E, Byrne MJ, Reeder P, Herskovic JR, Johnson CW, McCoy AB, et al. A benchmark comparison of deterministic and probabilistic methods for defining manual review datasets in duplicate records reconciliation. J Am Med Inform Assoc 2014;21(1):97-104 [FREE Full text] [doi: 10.1136/amiajnl-2013-001744] [Medline: 23703827]
22. Fatemi B, Kazemi SM, Poole D. Record linkage to match customer names: a probabilistic approach. arXiv. 2018. URL: http://arxiv.org/abs/1806.10928 [accessed 2021-11-24]
23. Kasiiti N, Wawira J, Purkayastha S, Were M. Comparative performance analysis of different fingerprint biometric scanners for patient matching. MEDINFO 2017: Precision Healthcare through Informatics. 2017. URL: https://scholarworks.iupui.edu/bitstream/handle/1805/16900/SHTI245-1053.pdf?sequence=1&isAllowed=y [accessed 2021-12-01]
24. Monica K. Patient matching strategy should leverage smartphones, biometrics. EHRIntelligence. 2018. URL: https://ehrintelligence.com/news/patient-matching-strategy-should-leverage-smartphones-biometrics [accessed 2021-11-25]
25. National Research Council. Biometric Recognition: Challenges and Opportunities. Washington, DC: The National Academies Press; 2010:1-182.
26. More hospitals using fingerprint scans for patient ID. Nairaland Forum. URL: https://www.nairaland.com/3209909/trust-hospital-scan-fingerprint-see [accessed 2021-11-25]
27. Blomeke CR, Elliott SJ, Walter TM. Bacterial survivability and transferability on biometric devices. In: Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology. 2007 Presented at: 41st Annual IEEE International Carnahan Conference on Security Technology; Oct. 8-11, 2007; Ottawa, ON, Canada. [doi: 10.1109/ccst.2007.4373472]
28. Chan S. New scanners for tracking city workers. New York Times. 2007. URL: https://www.nytimes.com/2007/01/23/nyregion/23scanning.html [accessed 2021-11-25]
29. World Health Organization. WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge Clean Care is Safer Care. Geneva: World Health Organization; 2009:1-270.
30. Chinedum AV. Design and implementation of a biometric information system. J Sci Res Innov Technol 2017;4(2):1-13 [FREE Full text]
31. Faridah Y, Nasir H, Kushsairy A, Safie SI, Khan S, Gunawan TS. Fingerprint biometric systems. Trends Bioinform 2016 Sep 15;9(2):52-58. [doi: 10.3923/tb.2016.52.58]
32. Bahaa-Eldin AM. A medium resolution fingerprint matching system. Ain Shams Eng J 2013 Sep;4(3):393-408. [doi: 10.1016/j.asej.2012.10.001]
33. Biometrics technology market size, share and trends analysis report, report by end-use (Government, Banking and Finance, Transport/Logistics, Defense and Security), by application (AFIS, Iris, Non-AFIS), and segment forecasts, 2018 - 2025. Grand View Research, Inc. 2018. URL: https://www.grandviewresearch.com/industry-analysis/biometrics-industry [accessed 2021-11-24]
34. Jhaveri H, Jhaveri H, Sanghavi D. Biometric security system and its applications in healthcare. Int J Tech Res Appl 2014;2(6):15-20 [FREE Full text]
35. Ali MM, Mahale VH, Yannawar P, Gaikwad AT. Overview of fingerprint recognition system. In: Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). 2016 Presented at: International

Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT); March 3-5, 2016; Chennai, India. [doi: 10.1109/iceeot.2016.7754900]
36. MacKinnon W, Wasserman M. Implementing electronic medical record systems. IT Prof 2009 Nov;11(6):50-53. [doi: 10.1109/mitp.2009.125]
37. OpenMRS - Medical Record System. URL: https://openmrs.org/ [accessed 2021-11-24]
38. Sadinle M, Fienberg SE. A generalized Fellegi–Sunter framework for multiple record linkage with application to homicide record systems. J Am Stat Assoc 2013 Jun;108(502):385-397 [FREE Full text] [doi: 10.1080/01621459.2012.757231]
39. Java Network Launch Protocol. Oracle. URL: https://docs.oracle.com/javase/tutorial/deployment/deploymentInDepth/jnlp.html [accessed 2021-11-24]
40. Maddala S, Tangellapally SR, Bartuněk JS, Nilsson M. Implementation and evaluation of NIST Biometric Image Software for fingerprint recognition. In: Proceedings of the ISSNIP Biosignals and Biorobotics Conference 2011. 2011 Presented at: ISSNIP Biosignals and Biorobotics Conference 2011; Jan. 6-8, 2011; Vitoria, Brazil. [doi: 10.1109/brc.2011.5740672]
41. Storisteanu DM, Norman TL, Grigore A, Norman TL. Biometric fingerprint system to enable rapid and accurate identification of beneficiaries. Glob Health Sci Pract 2015 Mar 05;3(1):135-137. [doi: 10.9745/ghsp-d-15-00010]
42. Odei-Lartey EO, Boateng D, Danso S, Kwarteng A, Abokyi L, Amenga-Etego S, et al. The application of a biometric identification technique for linking community and hospital data in rural Ghana. Glob Health Action 2016 Mar 17;9(1):29854 [FREE Full text] [doi: 10.3402/gha.v9.29854] [Medline: 26993473]
43. Shoewu O, Idowu OA. Development of attendance management system using biometrics. Pacific J Sci Technol 2012;13(1):1-8 [FREE Full text]
44. White EB, Meyer AJ, Ggita JM, Babirye D, Mark D, Ayakaka I, et al. Feasibility, acceptability, and adoption of fingerprint scanning during contact investigation for tuberculosis in Kampala, Uganda: a parallel-convergent, mixed-methods analysis. Peer J Preprints Preprint posted online on June 27, 2018. [doi: 10.7287/peerj.preprints.27007v1]
45. Grindrod K, Khan H, Hengartner U, Ong S, Logan AG, Vogel D, et al. Evaluating authentication options for mobile health applications in younger and older adults. PLoS One 2018 Jan 4;13(1):e0189048 [FREE Full text] [doi: 10.1371/journal.pone.0189048] [Medline: 29300736]
46. O'Doherty KC, Christofides E, Yen J, Bentzen HB, Burke W, Hallowell N, et al. If you build it, they will come: unintended future uses of organised health data collections. BMC Med Ethics 2016 Sep 06;17(1):54 [FREE Full text] [doi: 10.1186/s12910-016-0137-x] [Medline: 27600117]
47. North-Samardzic A. Biometric technology and ethics: beyond security applications. J Bus Ethics 2019 Mar 08;167(3):433-450. [doi: 10.1007/s10551-019-04143-6]
48. Sprokkereef A, De Hert P. Ethical practice in the use of biometric indentifiers within the EU. Law Sci Policy 2007;3(2):177-201 [FREE Full text]
49. Koda Y, Higuchi T, Jain AK. Advances in capturing child fingerprints: a high resolution CMOS Image Sensor with SLDR method. In: Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG). 2016 Presented at: International Conference of the Biometrics Special Interest Group (BIOSIG); Sept. 21-23, 2016; Darmstadt, Germany. [doi: 10.1109/biosig.2016.7736909]
50. Breithaupt R, Sousedik C, Meissner S. Full fingerprint scanner using optical coherence tomography. In: Proceedings of the 3rd International Workshop on Biometrics and Forensics (IWBF 2015). 2015 Presented at: 3rd International Workshop on Biometrics and Forensics (IWBF 2015); March 3-4, 2015; Gjovik, Norway. [doi: 10.1109/iwbf.2015.7110228]

Abbreviations

AMPATH: Academic Model Providing Access to Healthcare
EMRS: electronic medical record system
FAR: false acceptance rate
FER: failure to enroll rate
FRR: false rejection rate
LMIC: low- and middle-income country

Edited by R Kukafka, G Eysenbach; submitted 19.03.21; peer-reviewed by T Hannan, S Pesälä; comments to author 10.05.21; revised version received 16.06.21; accepted 03.10.21; published 22.12.21.

Please cite as:
Jaafa NK, Mokaya B, Savai SM, Yeung A, Siika AM, Were M
Implementation of Fingerprint Technology for Unique Patient Matching and Identification at an HIV Care and Treatment Facility in Western Kenya: Cross-sectional Study
J Med Internet Res 2021;23(12):e28958
URL: https://www.jmir.org/2021/12/e28958
doi:10.2196/28958
PMID:34941557

Original Paper

Electronic Health Record Implementations and Insufficient Training Endanger Nurses' Well-being: Cross-sectional Survey Study

Tarja Heponiemi[1], PhD; Kia Gluschkoff[1], PhD; Tuulikki Vehko[1], PhD; Anu-Marja Kaihlanen[1], PhD; Kaija Saranto[2], PhD; Sari Nissinen[2], PhD; Janna Nadav[1], MSc; Sari Kujala[3], PhD

[1]Finnish Institute for Health and Welfare, Helsinki, Finland

[2]University of Eastern Finland, Kuopio, Finland

[3]Aalto University, Espoo, Finland

Corresponding Author:
Tarja Heponiemi, PhD
Finnish Institute for Health and Welfare
Mannerheimintie 166
Helsinki, 00271
Finland
Phone: 358 295247434
Email: tarja.heponiemi@thl.fi

Abstract

Background: High expectations have been set for the implementations of health information systems (HIS) in health care. However, nurses have been dissatisfied after implementations of HIS. In particular, poorly functioning electronic health records (EHRs) have been found to induce stress and cognitive workload. Moreover, the need to learn new systems may require considerable effort from nurses. Thus, EHR implementations may have an effect on the well-being of nurses.

Objective: This study aimed to examine the associations of EHR-to-EHR implementations and the sufficiency of related training with perceived stress related to information systems (SRIS), time pressure, and cognitive failures among registered nurses. Moreover, we examined the moderating effect of the employment sector (hospital, primary care, social services, and others) on these associations.

Methods: This study was a cross-sectional survey study of 3610 registered Finnish nurses in 2020. EHR implementation was measured by assessing whether the work unit of each respondent had implemented or will implement a new EHR (1) within the last 6 months, (2) within the last 12 months, (3) in the next 12 months, and (4) at no point within the last 12 months or in the forthcoming 12 months. The associations were examined using analyses of covariance adjusted for age, gender, and employment sector.

Results: The highest levels of SRIS (adjusted mean 4.07, SE 0.05) and time pressure (adjusted mean 4.55, SE 0.06) were observed among those who had experienced an EHR implementation within the last 6 months. The lowest levels of SRIS (adjusted mean 3.26, SE 0.04), time pressure (adjusted mean 4.41, SE 0.05), and cognitive failures (adjusted mean 1.84, SE 0.02) were observed among those who did not experience any completed or forthcoming implementations within 12 months. Nurses who perceived that they had received sufficient implementation-related training experienced less SRIS (F_1=153.40, P<.001), time pressure (F_1=80.95, P<.001), and cognitive failures (F_1=34.96, P<.001) than those who had received insufficient training. Recent implementations and insufficient training were especially strongly associated with high levels of SRIS in hospitals.

Conclusions: EHR implementations and insufficient training related to these implementations may endanger the well-being of nurses and even lead to errors. Thus, it is extremely important for organizations to offer comprehensive training before, during, and after implementations. Moreover, easy-to-use systems that allow transition periods, a re-engineering approach, and user involvement may be beneficial to nurses in the implementation process. Training and other improvements would be especially important in hospitals.

(J Med Internet Res 2021;23(12):e27096) doi:10.2196/27096

KEYWORDS

electronic health records; implementation; information systems; training; stress; cognitive failures; time pressure; registered nurses

Introduction

Registered nurses form the largest group using health information systems (HIS) in health care. Therefore, the successful implementation of new HIS highly depends on nurses. High expectations have been set for the implementation of HIS in health care, for example, regarding increased efficiency. Indeed, a previous study shows that the implementation of intensive care unit information systems decreased the time nurses spent on documentation by over 30% and increased the time spent on direct patient care [1]. However, the outcomes of implementations may not always meet the high expectations [2]. Failure to understand users' needs and support workflow are some reasons why implementations may fail [3].

Registered nurses' work is often stressful and cognitively burdensome, and difficulties with HIS may induce extra stress and time pressure. Nurses have been found to prefer electronic health records (EHRs) over paper charts and think that EHR usage enhances nursing work but increases demand on work time and decreases the quality of care [4]. Among nurses, difficult-to-use EHRs have been associated with high time pressure and distress [5], and cognitive workload [6,7], which in turn have been associated with cognitive failures as well as lower patient safety [8].

Even though registered nurses are competent users of EHRs [9], the new systems that are implemented may be difficult and complicated to use especially in the beginning, thus becoming additionally demanding. This may lead to information chaos [10], which has been shown to result in decision-making errors and increased mental workload [10,11]. A previous study showed substantially increased cognitive workload related to new EHR implementation for nurses [6]. Nurses have been found to be most dissatisfied approximately 9 months after implementation, whereas their perceptions appeared to be more balanced after 18 months [12]. In another study, nurses reported greater acceptance of the EHR 12 months after implementation than after 3 months [13].

Education and training are essential prerequisites for successful HIS implementation [14]. The intricacies of new systems and changing functionalities may require nurses to allocate time and effort if they wish to master the new systems. However, the demands and pressures of care may not always afford nurses time to learn the new systems, and it is possible that not enough training and time have been allocated to this process [15,16]. Therefore, it may be a burden for nurses being forced to learn how to use the new systems effectively and efficiently, especially if they are not offered sufficient implementation-related training. Indeed, it has been found that HIS implementation may increase nurses' workload if they receive insufficient training before implementation [14]. Training has even been shown to have a positive impact on the perceived work environment as well [17].

Stress is an ambiguous concept with many definitions; for example, it has been defined as a relationship between a person and the environment that is appraised as important for an individual and exceeds coping resources [18]. Poorly functioning and constantly changing information systems may elicit this kind of stress appraisal, which can be designated as stress related to information systems (SRIS). For example, information systems have emerged as one of the highest stress-inducing factors among Finnish physicians alongside time pressure and patient-related stress [19-21]. Previous findings show that SRIS has increased in the 21st century among physicians [19,21], and the usability of EHRs has an effect on its levels [22]. However, SRIS is less studied among nurses and more information is needed. For example, nurses themselves have proposed that training could be one way to reduce nurses' SRIS levels [23].

As mentioned previously, studies show that nurses are dissatisfied after HIS implementations and challenges with HIS are associated with stress, time pressure, and cognitive burden [5-7,12,13,24]. Previous implementation studies have often focused on implementations involving transitions from paper-based systems to EHR systems [6,12]. However, many developed countries have already reached or will soon reach a saturation point where almost all health care organizations use EHRs. For example, in Finland, the EHR coverage has reached a saturation point of 100%, and many different brands of systems are in use [25]. Thus, more information is needed, especially on the effects of transitioning from one EHR system to another and implementation-related training on the stress levels and well-being of nurses. Therefore, our interest was on implementation of new brands of EHR systems, that is, the changes experienced when transitioning from one EHR to another.

Finnish nurses use many different EHR system brands; for example, in public hospitals, the 7 most popular brands are used by approximately 92% of nurses, and in primary care, the 4 most popular brands are used by 94% of nurses [26]. The most commonly used EHRs among nurses in Finland are Lifecare, Uranus, Pegasos, Apotti (system brand: Epic), Effica Healthcare, Mediatri, Esko, DynamicHealth, and DomaCare [26]. For example, large-scale implementations of Apotti were in progress in different areas of Helsinki and the Uusimaa region from 2018 to 2020.

This study examined the associations between EHR implementations and the sufficiency of training related to implementations with perceived SRIS, time pressure, and cognitive failures among Finnish registered nurses. Moreover, previous studies show that the employment sector, such as whether a person is employed in a hospital or a primary care center, plays an important role in health professionals' perceptions of EHRs and how they affect professionals' stress levels and well-being [22,27,28]. Therefore, we additionally examined whether the employment sector would have an effect on these associations.

Methods

Sample

The data were collected during the spring of 2020 through an internet-based Webropol survey. The link to the survey was sent via email by the Finnish Nurses Association, Tehy (The union of health and social care professionals in Finland) and the National Professional Association for the interests of experts

and managers in health care (TAJA) to their members under 65 years of age, including 58,276 nurses, midwives, and public health nurses representing 72% of the eligible population [29]. One reminder was sent to those who did not respond. A more detailed description of the study can be found elsewhere [29]. Altogether, 10,094 registered nurses opened the link and 3912 responded. Of those who responded, 302 answered that they did not perceive themselves as fit to answer the questionnaire because they had not worked as registered nurses for a long time. Thus, the final sample included 3610 respondents (93.1% women) aged between 22 and 65 years (mean 45.7, SD 11.0) [29]. The sample was representative of the eligible population in terms of the regionality and employment sector. Women were slightly overrepresented, and those under 40 years of age were slightly underrepresented [29]. According to a power analysis conducted using WebPower, an internet-based tool [30], the study had more than 95% power to detect small effects (f=0.1) with an α level of .05 in a 2×4 analysis of variance (ANOVA). Ethical approval for the study was provided by The Finnish Institute for Health and Welfare (THL/482/6.02.01/2020).

Measures

The questionnaire items used in the present study can be found in Multimedia Appendix 1.

Dependent Variables

SRIS was measured with the mean of 2 items, framed in 1 question that asked how often (during the last 6 months) the respondent had been distracted by, worried about, or stressed about (1) constantly changing HIS and (2) difficult, poorly performing information technology (IT) equipment or software. The answers were rated on a 6-point scale ranging from 1 (never) to 6 (constantly). The scale's reliability (Cronbach α=.74) was established in the present sample. This measure was developed in Finland when examining the health and well-being of physicians. [19-21]. It has previously been associated with, for example, experience in using information systems, cognitive workload, distress, and EHR usability [19,22].

Time pressure was measured with the mean of 2 items (α=.94) measuring how often (during the prior half-year period) a person had been distracted by, worried about, or stressed about (1) constantly being in a hurry and time pressure coming from unfinished work tasks and (2) having too little time to do work properly. The items were rated on a 6-point scale ranging from 1 (never) to 6 (constantly). This measure has been widely used previously and associated, for example, with the nurses' perceptions on the poor usability of EHRs [5].

Cognitive failures can be defined as "cognitively based errors that occur during the performance of a task that a person is normally successful in executing" [31]. They were measured with 3 items (α=.6) derived from the Workplace Cognitive Failure Scale (WCFS) [32,33]. Our survey included 1 item from each of the 3 dimensions of the WCFS: failure of memory, failure of attention, and failure of action. The chosen items have previously shown the highest loadings for their dimensions [33]. Participants were asked to rate how often they have faced situations at work where they (1) have not been able to remember work-related passwords, sets of numbers, etc. (memory failure); (2) have not fully listened to the instructions or requests they have received (attention failure); and (3) have accidentally started or closed the wrong device, system, or program (action failure). Items were rated on a 5-point scale ranging from 1 (never) to 5 (several times a day).

Independent Variables

EHR implementation was measured with a question asking whether the respondent's work unit had implemented or will implement a new EHR. The response options were (1) yes, within the last 6 months, (2) yes, within last 12 months, (3) no, but within the forthcoming 12 months, and (4) no past or forthcoming implementations within 12 months.

Training was assessed with a question asking whether the respondent had received sufficient training related to the required changes in work practices (such as new electronic documentation and care practices) due to HIS implementations. The answer options ranged between 1 (completely disagree) and 5 (completely agree). This question also included the answer option "cannot answer," which was coded as missing. The responses were coded as 0=insufficient training (answer options 1-3) and 1=sufficient training (answer options 4-5).

As control variables, gender, age, and employment sectors were also included in the survey. Employment sectors were coded as 1=hospital, 2=primary care, 3=social services, and 4=other.

Statistical Analysis

The associations of the implementation phase and training with the dependent variables were analyzed with analyses of covariance (in separate analyses for each dependent variable). The analyses were adjusted for age, gender, and employment sector. The interactions of the employment sector with the implementation phase and training for the dependent variables were examined with analyses of covariance adjusted for age, gender, and primary effects (in separate analyses for each interaction and dependent variable). Respondents who had missing data for a given variable were excluded from the analyses of that variable. Thus, due to missing information in some variables, n varied between 3525 and 3610.

To further examine the validity of our dependent variables, we conducted principal components analysis (PCA) through direct oblimin rotation with the items of the dependent variables (SRIS, time pressure, and cognitive failures). Moreover, the analyses of covariance were repeated for analyzing sensitivity using the principal component scores resulting from these analyses as the dependent variables.

Results

Demographics

The characteristics of the study population are given in Table 1. Approximately 25% (834/3610) of the respondents had experienced EHR implementation within the preceding 6 months, 13% (476/3610) within the preceding 12 months, and 20% (714/3610) reported forthcoming EHR implementation within 12 months. More than half of the respondents (1894/3573) reported insufficient training regarding changes required in the way of working due to HIS implementations.

Table 1. Basic background characteristics of the study sample (N=3610[a]).

Characteristic	Value
Gender, n (%)	
Women	3340 (93.1)
Men	249 (6.9)
Employment sector, n (%)	
Hospital	1903 (52.7)
Primary care	795 (22)
Social services	445 (12.3)
Other	467 (12.9)
EHR[b] implementation phase, n (%)	
Yes, within the last 6 months	834 (23.1)
Yes, within the last 12 months	476 (13.2)
No, but forthcoming within the next 12 months	714 (19.8)
No	1586 (43.9)
Training related to implementation, n (%)	
Insufficient	1894 (53)
Sufficient	1679 (47)
Age,[c] mean (SD)	45.68 (10.97)
Stress related to information systems,[c] mean (SD)	3.7 (1.13)
Time pressure,[d] mean (SD)	4.54 (1.12)
Cognitive failures,[e] mean (SD)	1.88 (0.5)

[a]Due to missing information in some variables, n varies between 3573 and 3610.

[b]EHR: electronic health record.

[c]Ranged between 22 and 67.

[d]Ranged between 1 and 6.

[e]Ranged between 1 and 5.

Main Effects for SRIS

Table 2 shows the results of analyses of covariance. Age, gender, employment sector, implementation phase, and training were all associated with SRIS. Women had higher levels of SRIS than men. Higher age was associated with higher levels of SRIS. The highest level of SRIS was in hospitals and the lowest was in social care. As observed in Table 3, those who had experienced EHR implementation within the preceding 6 months perceived the highest levels of SRIS, whereas those who did not have to experience forthcoming or prior implementations within 12 months perceived the lowest levels of SRIS. Those who perceived that they had received sufficient training had lower levels of SRIS compared to those who perceived that they had not received sufficient training.

Table 2. Associations among explanatory factors with stress related to information systems, time pressure, and cognitive failures (analysis of covariance[a]).

Variable	SRIS[b]		Time pressure		Cognitive failures	
	F test *(df)*	*P* value	*F* test *(df)*	*P* value	*F* test *(df)*	*P* value
Age	15.47 (1)	<.001	30.31 (1)	<.001	4.96 (1)	.03
Gender	9.50 (1)	.003	20.57 (1)	<.001	0.21 (1)	.65
Sector	24.14 (3)	<.001	1.77 (3)	.23	1.36 (3)	.25
Implementation	118.43 (3)	<.001	3.05 (3)	.03	5.36 (3)	.001
Training	153.40 (1)	<.001	80.95 (1)	<.001	34.96 (1)	<.001
R^2	0.165	N/A[c]	0.033	N/A	0.015	N/A

[a]Due to missing information in some variables, n varies between 3525 and 3546.

[b]SRIS: stress related to information systems.

[c]N/A: not applicable.

Table 3. Estimated marginal means[a] of stress related to information systems, time pressure, and cognitive failures according to implementation phase.[b]

Implementation phase	SRIS[c]		Time pressure		Cognitive failures	
	Mean	SE	Mean	SE	Mean	SE
Within the last 6 months	4.07	0.05	4.55	0.06	1.87	0.03
Within the last 12 months	3.76	0.06	4.45	0.06	1.94	0.03
Forthcoming within the next 12 months	3.44	0.05	4.45	0.06	1.86	0.03
No implementations	3.26	0.04	4.41	0.05	1.84	0.02

[a]Adjusted for age, gender, employment sector, and training.

[b]Due to missing information in some variables, n varies between 3525 and 3546.

[c]SRIS: stress related to information systems.

Main Effects for Time Pressure

Age, gender, implementation phase, and training were associated with time pressure. Women had higher levels of time pressure than men. Higher age was associated with lower levels of time pressure. Those who had experienced EHR implementations within the preceding 6 months had the highest levels of time pressure, whereas those who did not have to experience forthcoming implementations or postimplementation outcomes within 12 months had the lowest levels of time pressure, as indicated in Table 3. Sufficient training was associated with low levels of time pressure.

Main Effects for Cognitive Failures

Age, implementation phase, and training were associated with cognitive failures. Higher age was associated with lower levels of cognitive failures. Those who had experienced EHR implementation within the preceding 12 months had higher levels of cognitive failures compared to other groups (Table 3). Those who perceived that they had undergone sufficient training had lower levels of cognitive failures compared to those who perceived that they had not had sufficient training.

Interactions With the Employment Sector

The interaction between the implementation phase and employment sector was significant for SRIS (F_9=4.32, P<.001). As observed in Figure 1, implementation in hospitals within 6 months is associated with higher SRIS levels than other sectors (n of the different groups varied between 49 and 715). Moreover, the interaction between training and the employment sector was significant for SRIS (F_3=10.18, P<.001). SRIS levels in hospitals were particularly high if insufficient training was perceived, as shown in Figure 2.

The interaction between training and the employment sector was significant for time pressure (F_3=4.18, P=.006). In primary care, sufficient training was not so strongly associated with time pressure, whereas in all other sectors, time pressure levels were low if training was perceived to be sufficient, as observed in Figure 3. The interaction between the implementation phase and employment sector was not significant for time pressure.

The interactions of the employment sector with the implementation phase and training were not significant for cognitive failures.

Figure 1. Interaction between implementation phase and employment sectors for stress related to information systems (*P*<.001). SRIS: stress related to information systems.

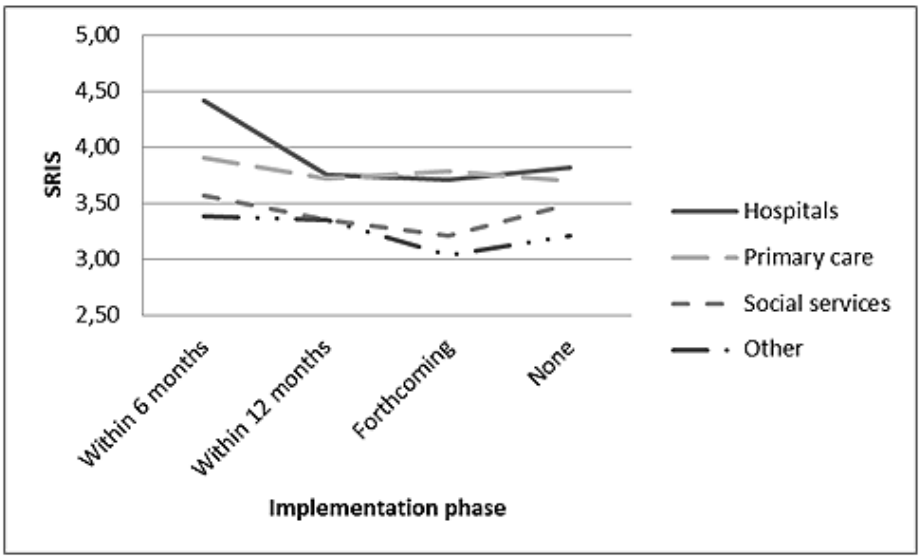


Figure 2. Interaction between training and employment sectors for stress related to information systems (*P*<.001). SRIS: stress related to information systems.

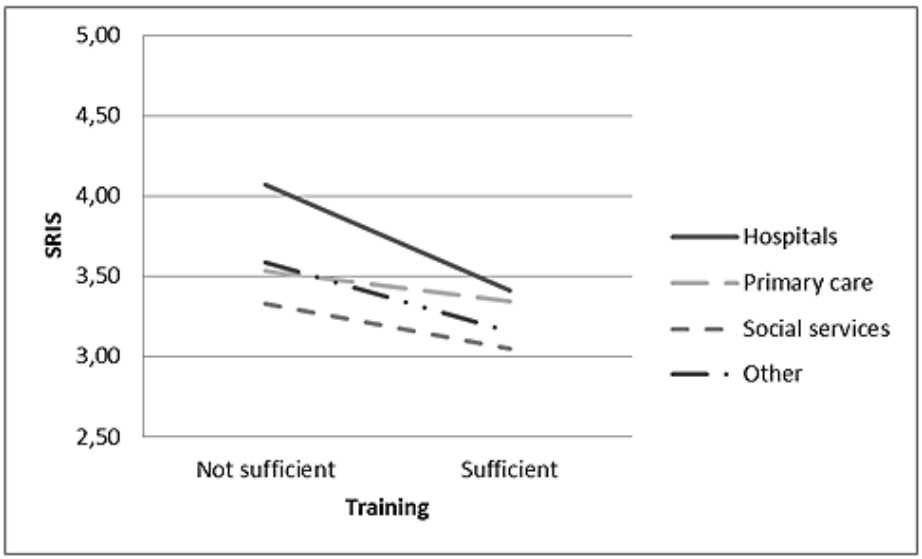
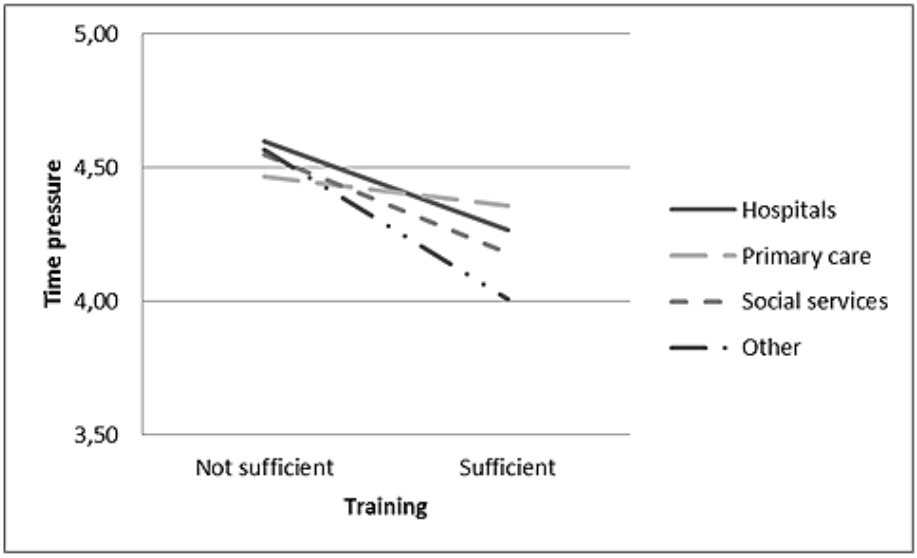


Figure 3. Interaction between training and employment sectors for time pressure (*P*=.006).



Sensitivity Analyses

The results of PCA are given in Multimedia Appendix 2. Results showed that the principal components resulting from the analysis were similar, as the original measures used (SRIS, time pressure, and cognitive failures) and the items loaded well with these principal components. ANOVA conducted with the principal component scores resulting from PCA as the dependent variables showed results corresponding with the analyses conducted using original variables (see Multimedia Appendix 3). Moreover, the interaction results with the principal components scores were congruent with the results from the original analyses.

Discussion

Principal Results

Our results show that EHR implementations have the potential to decrease the well-being of registered nurses. More specifically, we found that the highest levels of stress related to poorly functioning information systems and time pressure were experienced among those who had experienced EHR implementation within the preceding 6 months. In particular, recent implementations were strongly associated with high levels of SRIS in hospitals. The highest levels of cognitive failures were instead experienced among those who had experienced EHR implementation within the preceding 6 to 12 months. The lowest levels of SRIS, time pressure, and cognitive failures were experienced among those who did not have any past or forthcoming implementations within 12 months.

Sufficient training related to implementations appears extremely crucial for nurses and is associated with improved well-being. Those nurses who perceived that they had received sufficient training related to the changes required in work practices due to HIS implementations experienced less SRIS, time pressure, and cognitive failures. The highest levels of SRIS were among those nurses who worked in hospitals and did not receive sufficient training.

Limitations

We used self-reported measures, which always poses a question related to problems associated with an inflation of the strengths of relationships and common method variance. The reliabilities of our scales were good, except that the reliability of the cognitive failures scale was 0.6, which can be considered low. However, this reliability can still be considered acceptable because the scale included only 3 items [34]. Moreover, we used cross-sectional survey data; thus, causal inferences cannot be drawn from our results. It is possible that nurses who are more competent in using EHRs also learn to use the new system more quickly and are also more likely to perceive that they have received sufficient training.

In addition, a major limitation of this study is the low explanatory strength of the independent variables used for time pressure and the cognitive failures variables, which can be seen in the low R^2 values of the analyses of covariance regarding these variables. Consequently, our results should be interpreted with caution and future studies with high-quality validated measures are still needed on this subject. However, even the small effect size may be noteworthy, and it has been suggested that even though the size of the effect in psychological research would be very small, it may potentially be very consequential in the long run [35]. The effect size was larger when explaining the SRIS variable, which is logical given that when a new EHR is implemented, the strain and stress observed among nurses are not surprising.

Moreover, although we adjusted our analyses for age, gender, and employment sectors, the possibility of residual confounding cannot be totally eliminated. For example, it is possible that some unknown third variable may have an effect on the stress and perceived training level and consequently explain the relationship between training and our dependent variables.

Finland is a country with universal health care for all residents and one of the forerunners in the digitalization of health care [36]. Therefore, we must be cautious in generalizing our findings to countries with dissimilar health care systems or HIS.

We used a rather large sample of registered nurses (3610 nurses), which was obtained from the registers of academic associations and trade unions and may have affected the representativeness of the sample. The email invitation to participate in the survey was sent to 72% of the eligible population and our sample represented the eligible population in terms of the regionality and employment sector but included a slightly disproportionate cohort of women and those over 40 years of age [29]. The data were collected in the spring of 2020 (March to April) at the time when the COVID-19 pandemic gained prominence in Finland. The most stringent restrictions so far were implemented in the middle of March 2020. Therefore, only 1 reminder was sent to those who had not responded to the invite. These circumstances may have had an effect on the results, especially in those hospitals that were most strongly affected by the pandemic.

Comparison With Prior Work

Our results show that EHR implementations may endanger the well-being of nurses. This is congruent with previous findings showing that EHR implementation is associated with decreased interdisciplinary communication, a high demand on work time, and low perceived quality of care among nurses [4]. Correspondingly, nurses have been found to experience stress due to added work, along with concerns about security and encountering poor cooperation in the early stage after the implementation of the nursing information system [15]. It has also been shown previously that implementation seems to induce stress, frustration, and feelings of incompetency, especially among those nurses who have problems with tasks requiring digital skills [37].

According to our results, recent implementations occurring especially within the preceding 6 months seem to induce SRIS and time pressure. Previous studies have also obtained congruent findings [6,12,13]. The dissatisfaction seems to be the greatest soon after implementation and then declines, moving toward greater acceptance 12 months after implementation [12,13]. However, we found that cognitive failures were the highest from 6 to 12 months after implementation. A previous study has found that cognitive workload among nurses is the highest just after the EHR implementation and then returns toward baseline after 4 months [6]. There are many possible reasons for our finding that cognitive failures were the highest 6 to 12 months after implementation. For example, it may pertain to implementation-related support from vendors and organizations. It is also possible that nurses are protected from other cognitively burdensome tasks immediately after implementation.

Our findings show that besides implementation aspects related to the change from paper-based documentation to EHRs, implementation factors regarding the change from one EHR to another EHR may affect employees' well-being. Traditionally, the focus of previous studies has been on the effects of transitioning from paper-based documentation to EHRs [4,6,12]; however, studies focusing on the effects of transitioning from one EHR to another are emerging [13]. In future, we may expect research findings focusing on the transition from one EHR to another, given the widespread use of EHRs in developed countries.

Our results suggest that organizations should implement measures to decrease the negative impact of implementations. A meta-analysis suggested a re-engineering approach to better integrate HIS implementation in health care workflows [3]. During re-engineering, organizations should examine and consider restructuring their work processes related to operational factors and infrastructure in a manner that could enable them to optimally use HIS functions. Moreover, improving the usability of systems would support the implementation, decrease the need for training, and improve employee well-being [22,38,39]. The users should be allowed a transition period, giving them time to understand and appreciate the outcome of the system implementation [3]. Further, user involvement, strong leadership, project management techniques, and standards are important in ensuring successful implementation [3,39,40]. To improve the experience of nurses in the beginning stage, it would be important to commit the nurses to the system design early on [15].

Our findings highlight the importance of proper training related to implementations. According to our findings, it may be possible to decrease the negative ramifications of implementations on nurses' well-being and cognitive functioning with sufficient training. However, training is insufficient in many cases. In our study, 53% reported that training was insufficient. In another study, 62% of health care staff reported that they had not received enough training related to inpatient portal implementation [16]. Moreover, nurses have indicated in focus groups that they had insufficient training related to the nursing information systems [15].

Our results showed some variations according to the employment sectors. In hospitals, the stress levels associated with implementations were the highest and the training related to systems was especially important. This finding effectively reflects the previous findings among physicians showing that attitudes toward EHRs are most critical in hospitals [41-43]. It is also possible that hospitals especially experience insufficiencies in IT support, given that nurses in hospitals often also work outside of office hours. A previous study has suggested that in addition to training, organizations should also identify and appoint champions who could learn more thoroughly and teach others how to use different systems [44].

In primary care, it seems that sufficient training is also inadequate to buffer against time pressure. In Finnish primary care, one of the reasons for time pressure among nurses is that primary care involves accessibility problems and long waiting times [45]. A previous study among Finnish nurses showed that technical problems and poor user-friendliness of the EHRs are associated with high time pressure [5]. Thus, in primary care, tackling the problems associated with the usability of the systems that are implemented might be important in terms of time pressure.

Our results show that sufficient training related to implementations is highly important. It might be beneficial to offer training and other support during the whole implementation

period, that is, before, during, and after the implementation [46,47]. The amount of education has been found to be positively correlated with nurses 'attitudes and behaviors toward the implemented IS; thus, it is important that organizations provide quickly and easily accessible in-house support and proactive training in the use of HIS [48]. In addition, evidence indicates that it is important to consider which training methods would best support professionals in developing the necessary skills and using the systems. Adequate education and training encourage employees to use HIS, which are prerequisites for benefiting from implementations [3]. According to our results, training related to changing work practices due to HIS implementations is particularly important. Additionally, training featuring improvements such as keyboard entry skills, redesigning workflow, and improving interdisciplinary communication are considered necessary [15].

We suggest that training should be planned carefully in advance, including basic training at least 2 to 3 weeks before implementation. After implementation, training should continue for several weeks, following which the authorities must assess whether more training is needed. Moreover, it is important to provide time for nurses to learn to use the systems in practice.

Conclusions

The present study shows that EHR-to-EHR implementations and insufficient training related to the implementations may impair nurses' well-being and even lead to cognitive failures. Thus, it is crucial that organizations implement measures to decrease the negative ramifications of implementations on their employees. This would be very important in all sectors, but especially in hospitals.

Acknowledgments

This study was supported by the Ministry of Social Affairs and Health (project 414919001) and the Strategic Research Council at the Academy of Finland (projects 327145 and 327147). None of them had any role in the design of the study and collection, analysis, and interpretation of the data or in the writing of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measures used in the study.
[DOCX File , 20 KB - jmir_v23i12e27096_app1.docx]

Multimedia Appendix 2
Results of principal component analysis with direct oblimin rotation: component loadings and variance explained.
[DOCX File , 14 KB - jmir_v23i12e27096_app2.docx]

Multimedia Appendix 3
Association of explanatory factors with principal component scores for stress related to information systems, time pressure, and cognitive failures (analysis of covariance).
[DOCX File , 14 KB - jmir_v23i12e27096_app3.docx]

References

1. Wong DH, Gallegos Y, Weinger MB, Clack S, Slagle J, Anderson CT. Changes in intensive care unit nurse task activity after installation of a third-generation intensive care unit information system. Crit Care Med 2003 Oct;31(10):2488-2494. [doi: 10.1097/01.CCM.0000089637.53301.EF] [Medline: 14530756]
2. Munyisia EN, Yu P, Hailey D. Does the introduction of an electronic nursing documentation system in a nursing home reduce time on documentation for the nursing staff? Int J Med Inform 2011 Nov;80(11):782-792. [doi: 10.1016/j.ijmedinf.2011.08.009] [Medline: 21956002]
3. Rahimi B, Vimarlund V, Timpka T. Health information system implementation: a qualitative meta-analysis. J Med Syst 2009 Oct;33(5):359-368. [doi: 10.1007/s10916-008-9198-9] [Medline: 19827262]
4. Kossman SP, Scheidenhelm SL. Nurses' perceptions of the impact of electronic health records on work and patient outcomes. Comput Inform Nurs 2008 Mar;26(2):69-77. [doi: 10.1097/01.NCN.0000304775.40531.67] [Medline: 18317257]
5. Vehko T, Hyppönen H, Puttonen S, Kujala S, Ketola E, Tuukkanen J, et al. Experienced time pressure and stress: electronic health records usability and information technology competence play a role. BMC Med Inform Decis Mak 2019 Aug;19(1):160 [FREE Full text] [doi: 10.1186/s12911-019-0891-z] [Medline: 31412859]
6. Colligan L, Potts HWW, Finn CT, Sinkin RA. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. Int J Med Inform 2015 Jul;84(7):469-476. [doi: 10.1016/j.ijmedinf.2015.03.003] [Medline: 25868807]

7. Saitwal H, Feng X, Walji M, Patel V, Zhang J. Assessing performance of an electronic health record (EHR) using cognitive task analysis. Int J Med Inform 2010 Jul;79(7):501-506. [doi: 10.1016/j.ijmedinf.2010.04.001] [Medline: 20452274]
8. Mohammady MJ, Sedighi A, Khaleghdoost T, Nejad EK, Javadi-Pashaki N. Impacts of occupational cognitive failure and subjective workload on patient safety incidents among intensive care units nurses. Indian J Occup Environ Med 2020;24(2):96-101 [FREE Full text] [doi: 10.4103/ijoem.IJOEM_22_20] [Medline: 33281379]
9. Kinnunen U, Heponiemi T, Rajalahti E, Ahonen O, Korhonen T, Hyppönen H. Factors related to health informatics competencies for nurses-results of a national electronic health record survey. Comput Inform Nurs 2019 Aug;37(8):420-429. [doi: 10.1097/CIN.0000000000000511] [Medline: 30741730]
10. Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751 [FREE Full text] [doi: 10.3122/jabfm.2011.06.100255] [Medline: 22086819]
11. Einhorn HJ. Learning from experience and suboptimal rules in decision making. In: Kahneman D, Tversk Y, Slovic P, editors. Judgement under Uncertainty: Heuristics & Biases. Cambridge, UK: Cambridge University Press; 1982:268-287.
12. Zadvinskis IM, Garvey Smith J, Yen P. Nurses' experience with health information technology: longitudinal qualitative study. JMIR Med Inform 2018 Jun;6(2):e38 [FREE Full text] [doi: 10.2196/medinform.8734] [Medline: 29945862]
13. Carayon P, Cartmill R, Blosky MA, Brown R, Hackenberg M, Hoonakker P, et al. ICU nurses' acceptance of electronic health records. J Am Med Inform Assoc 2011 Nov;18(6):812-819. [doi: 10.1136/amiajnl-2010-000018]
14. Littlejohns P, Wyatt J, Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. BMJ 2003 Apr;326(7394):860-863 [FREE Full text] [doi: 10.1136/bmj.326.7394.860] [Medline: 12702622]
15. Lee T. Nurses' experiences using a nursing information system: early stage of technology implementation. Comput Inform Nurs 2007;25(5):294-300. [doi: 10.1097/01.NCN.0000289166.61863.0b] [Medline: 17827993]
16. Hefner JL, Sieck CJ, Walker DM, Huerta TR, McAlearney AS. System-wide inpatient portal implementation: survey of health care team perceptions. JMIR Med Inform 2017 Sep;5(3):e31 [FREE Full text] [doi: 10.2196/medinform.7707] [Medline: 28912115]
17. De Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020 Aug;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]
18. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984.
19. Heponiemi T, Hyppönen H, Kujala S, Aalto A, Vehko T, Vänskä J, et al. Predictors of physicians' stress related to information systems: a nine-year follow-up survey study. BMC Health Serv Res 2018 Apr;18(1):284 [FREE Full text] [doi: 10.1186/s12913-018-3094-x] [Medline: 29653530]
20. Heponiemi T, Vänskä J, Aalto A, Elovainio M. Potilastyöhön ja tietojärjestelmiin liittyvä stressi lisääntyi: Kyselyt lääkäreille 2006-2010 changes in stress factors and psychosocial resources among Finnish physicians between 2006 and 2010. Finn Med J 2012;47:3491-3495.
21. Heponiemi T, Hyppönen H, Vehko T, Kujala S, Aalto A, Vänskä J, et al. Finnish physicians' stress related to information systems keeps increasing: a longitudinal three-wave survey study. BMC Med Inform Decis Mak 2017 Oct;17(1):147 [FREE Full text] [doi: 10.1186/s12911-017-0545-y] [Medline: 29041971]
22. Heponiemi T, Kujala S, Vainiomäki S, Vehko T, Lääveri T, Vänskä J, et al. Usability factors associated with physicians' distress and information system-related stress: cross-sectional survey. JMIR Med Inform 2019 Nov;7(4):e13466 [FREE Full text] [doi: 10.2196/13466] [Medline: 31687938]
23. AlQahtani M, AlShaibani W, AlAmri E, Edward D, Khandekar R. Electronic health record-related stress among nurses: determinants and solutions. Telemed J E Health 2020 Aug;27(5):544-550. [doi: 10.1089/tmj.2020.0059] [Medline: 32857018]
24. Kaihlanen A, Gluschkoff K, Hyppönen H, Kaipio J, Puttonen S, Vehko T, et al. The associations of electronic health record usability and user age with stress and cognitive failures among Finnish registered nurses: cross-sectional study. JMIR Med Inform 2020 Nov;8(11):e23623 [FREE Full text] [doi: 10.2196/23623] [Medline: 33206050]
25. Reponen J, Kangas M, Hämäläinen P, Haverinen J, Keränen N. Availability and use of e-health in Finland. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-health and e-welfare of Finland: Check Point 2018. Helsinki, Finland: National Institute for Health and Welfare (THL); 2019:52-84.
26. Kyytsönen M, Hyppönen H, Koponen S, Kinnunen U, Saranto K, Kivekäs E, et al. Tietojärjestelmät sairaanhoitajien työn tukena eri toimintaympäristöissä: kokemuksia tuotemerkeittäin. FinJeHeW 2020 Oct;12(3):250-269. [doi: 10.23996/fjhw.95704]
27. Kaipio J, Kuusisto A, Hyppönen H, Heponiemi T, Lääveri T. Physicians' and nurses' experiences on EHR usability: comparison between the professional groups by employment sector and system brand. Int J Med Inform 2020 Feb;134:104018 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104018] [Medline: 31835158]
28. Vainiomäki S, Heponiemi T, Vänskä J, Hyppönen H. Tailoring EHRs for specific working environments improves work well-being of physicians. Int J Environ Res Public Health 2020 Jun;17(13):4715 [FREE Full text] [doi: 10.3390/ijerph17134715] [Medline: 32630043]

29. Saranto K, Kinnunen U, Koponen S, Kyytsönen M, Hyyppönen H, Vehko T. Sairaanhoitajien valmiudet tiedonhallintaan sekä kokemukset potilas- ja asiakastietojärjestelmien tuesta työtehtäviin. FinJeHeW 2020 Oct;12(3):212-228 [FREE Full text] [doi: 10.23996/fjhw.95711]
30. Zhang Z, Yuan K. Practical Statistical Power Analysis Using WebPower and R. Granger, IN: ISDSA Press; 2018.
31. Martin M. Cognitive failure: everyday and laboratory performance. Bull Psychon Soc 2013 Oct;21(2):97-100. [doi: 10.3758/bf03329964]
32. Kalakoski V, Selinheimo S, Valtonen T, Turunen J, Käpykangas S, Ylisassi H, et al. Effects of a cognitive ergonomics workplace intervention (CogErg) on cognitive strain and well-being: a cluster-randomized controlled trial. a study protocol. BMC Psychol 2020 Jan;8(1):1 [FREE Full text] [doi: 10.1186/s40359-019-0349-1] [Medline: 31898551]
33. Wallace JC, Chen G. Development and validation of a work-specific measure of cognitive failure: implications for occupational safety. J Occup Organ Psychol 2005 Dec;78(4):615-632. [doi: 10.1348/096317905X37442]
34. Hair J, Black W, Babin B, Anderson R, Tatham R. Multivariate Data Analysis. New Jersey: Pearson Educational, Inc; 2006.
35. Funder DC, Ozer DJ. Evaluating effect size in psychological research: sense and nonsense. Adv Methods Pract Psychol Sci 2019 May;2(2):156-168. [doi: 10.1177/2515245919847202]
36. Sabes-Figuera R, Maghiros I. European Hospital Survey: Benchmarking Deployment of e-Health Services (2012-2013): Composite Indicators on eHealth Deployment and on Availability and Use of eHealth Functionalities. Luxembourg: Publications Office of the European Union; 2013.
37. De Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020 Aug;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]
38. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
39. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
40. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
41. Vänskä J, Viitanen J, Hyppönen H, Elovainio M, Winblad I, Reponen J. Lääkärien arviot potilastietojärjestelmistä kriittisiä. Finn Med J 2010;50:52-83.
42. Emani S, Ting DY, Healey M, Lipsitz SR, Karson AS, Einbinder JS, et al. Physician beliefs about the impact of meaningful use of the EHR: a cross-sectional study. Appl Clin Inform 2014;5(3):789-801 [FREE Full text] [doi: 10.4338/ACI-2014-05-RA-0050] [Medline: 25298817]
43. Winblad I, Hyppönen H, Vänskä J, Reponen J, Viitanen J, Elovainio M. Potilastietojärjestelmät tuotemerkittäin arvioitu: Kaikissa on kehitettävää. Finn Med J 2010;50:52-94.
44. Kujala S, Heponiemi T, Hilama P. Clinical leaders' self-perceived eHealth competencies in the implementation of new eHealth services. Stud Health Technol Inform 2019 Aug;264:1253-1257. [doi: 10.3233/SHTI190427] [Medline: 31438126]
45. Keskimäki I, Tynkkynen LK, Reissell E, Koivusalo M, Syrjä V, Vuorenkoski L, et al. Finland: health system review. Health Syst Transit 2019;21(2):1-166. [doi: 10.1007/978-1-349-95810-8_934]
46. Keshavjee K, Bosomworth J, Copen J, Lai J, Kucukyazici B, Lilani R, et al. Best practices in EMR implementation: a systematic review. AMIA Annu Symp Proc 2006:982 [FREE Full text] [Medline: 17238601]
47. Aguirre RR, Suarez O, Fuentes M, Sanchez-Gonzalez MA. Electronic health record implementation: a review of resources and tools. Cureus 2019 Sep;11(9):e5649 [FREE Full text] [doi: 10.7759/cureus.5649] [Medline: 31700751]
48. Ifinedo P. The moderating effects of demographic and individual characteristics on nurses' acceptance of information systems: a Canadian study. Int J Med Inform 2016 Mar;87:27-35. [doi: 10.1016/j.ijmedinf.2015.12.012] [Medline: 26806709]

Abbreviations

EHRs: electronic health records
HIS: health information systems
IT: information technology
PCA: principal components analysis
SRIS: stress related to information systems
WCFS: Workplace Cognitive Failure Scale

Edited by R Kukafka; submitted 11.01.21; peer-reviewed by Y Chu, M Esdar, A Fagerlund; comments to author 28.01.21; revised version received 22.02.21; accepted 10.11.21; published 23.12.21.

Please cite as:
Heponiemi T, Gluschkoff K, Vehko T, Kaihlanen AM, Saranto K, Nissinen S, Nadav J, Kujala S
Electronic Health Record Implementations and Insufficient Training Endanger Nurses' Well-being: Cross-sectional Survey Study
J Med Internet Res 2021;23(12):e27096
URL: https://www.jmir.org/2021/12/e27096
doi:10.2196/27096
PMID:34941546

Original Paper

Health Care Analytics With Time-Invariant and Time-Variant Feature Importance to Predict Hospital-Acquired Acute Kidney Injury: Observational Longitudinal Study

Horng-Ruey Chua[1,2], MMED; Kaiping Zheng[3], PhD; Anantharaman Vathsala[1,2], MD; Kee-Yuan Ngiam[4,5], MMED; Hui-Kim Yap[6,7], PhD; Liangjian Lu[6,7], MBBCHIR; Ho-Yee Tiong[5,8], MMED; Amartya Mukhopadhyay[2,9], MD; Graeme MacLaren[5,10], MSc; Shir-Lynn Lim[2,11], MMED; K Akalya[1], BSc; Beng-Chin Ooi[3], PhD

[1]Division of Nephrology, Department of Medicine, National University Hospital, Singapore, Singapore
[2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[3]Department of Computer Science, School of Computing, National University of Singapore, Singapore, Singapore
[4]Division of Endocrine Surgery, Department of Surgery, National University Hospital, Singapore, Singapore
[5]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[6]Division of Paediatric Nephrology, Department of Paediatrics, National University Children's Medical Institute, Singapore, Singapore
[7]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[8]Department of Urology, National University Hospital, Singapore, Singapore
[9]Division of Respiratory and Critical Care Medicine, Department of Medicine, National University Hospital, Singapore, Singapore
[10]Cardiothoracic Intensive Care Unit, Department of Cardiac, Thoracic and Vascular Surgery, National University Hospital, Singapore, Singapore
[11]Department of Cardiology, National University Heart Centre, Singapore, Singapore

Corresponding Author:
Horng-Ruey Chua, MMED
Division of Nephrology
Department of Medicine
National University Hospital
Level 10 Medicine, NUHS Tower Block
1E Kent Ridge Road
Singapore, 119228
Singapore
Phone: 65 67726178
Email: horng_ruey_chua@nuhs.edu.sg

Abstract

Background: Acute kidney injury (AKI) develops in 4% of hospitalized patients and is a marker of clinical deterioration and nephrotoxicity. AKI onset is highly variable in hospitals, which makes it difficult to time biomarker assessment in all patients for preemptive care.

Objective: The study sought to apply machine learning techniques to electronic health records and predict hospital-acquired AKI by a 48-hour lead time, with the aim to create an AKI surveillance algorithm that is deployable in real time.

Methods: The data were sourced from 20,732 case admissions in 16,288 patients over 1 year in our institution. We enhanced the bidirectional recurrent neural network model with a novel time-invariant and time-variant aggregated module to capture important clinical features temporal to AKI in every patient. Time-series features included laboratory parameters that preceded a 48-hour prediction window before AKI onset; the latter's corresponding reference was the final in-hospital serum creatinine performed in case admissions without AKI episodes.

Results: The cohort was of mean age 53 (SD 25) years, of whom 29%, 12%, 12%, and 53% had diabetes, ischemic heart disease, cancers, and baseline eGFR <90 mL/min/1.73 m^2, respectively. There were 911 AKI episodes in 869 patients. We derived and validated an algorithm in the testing dataset with an AUROC of 0.81 (0.78-0.85) for predicting AKI. At a 15% prediction threshold, our model generated 699 AKI alerts with 2 false positives for every true AKI and predicted 26% of AKIs. A lowered 5% prediction threshold improved the recall to 60% but generated 3746 AKI alerts with 6 false positives for every true AKI. Representative

interpretation results produced by our model alluded to the top-ranked features that predicted AKI that could be categorized in association with sepsis, acute coronary syndrome, nephrotoxicity, or multiorgan injury, specific to every case at risk.

Conclusions: We generated an accurate algorithm from electronic health records through machine learning that predicted AKI by a lead time of at least 48 hours. The prediction threshold could be adjusted during deployment to optimize recall and minimize alert fatigue, while its precision could potentially be augmented by targeted AKI biomarker assessment in the high-risk cohort identified.

(J Med Internet Res 2021;23(12):e30805) doi:10.2196/30805

KEYWORDS

acute kidney injury; artificial intelligence; biomarkers; clinical deterioration; electronic health records; hospital medicine; machine learning

Introduction

The clinical burden of acute kidney injury (AKI) worsens globally with the increasing complexity of cardiovascular diseases, anticancer therapy, and aging population [1-3]. AKI develops in 4% of patients admitted to our institution and involves more than 3000 patients annually [4]. A total of 39% of AKI cases develop during hospitalization following clinical deterioration and multiorgan dysfunction [4,5]. Additionally, 15% of patients who receive antimicrobials or chemotherapy of nephrotoxic potential develop drug-induced AKI [6,7]. Iodinated contrast administered for angiography contributes to AKI in 10% to 40% of patients with chronic kidney disease [8,9]. Once AKI develops in patients, however, the management remains supportive with control of its underlying triggers. AKI portends a poor patient prognosis with high mortality, prolonged hospitalization, and sustained deterioration of kidney function, with a significant risk of kidney failure in the long term [10,11].

Management strategies for high-risk patients may prevent AKI or reduce its downstream complications should AKI still develop. These measures must be implemented promptly, which requires the diagnosis of AKI in the subclinical phase, way before its onset. As the onset of AKI is highly variable during a patient's stay, it is unclear how best to time biomarker surveillance for kidney injury concerning the patient's clinical progress. The advent of electronic health records (EHRs) now provides us with real-time clinical data from routine patient care, built into millions of data points for analytics. These, along with AKI being defined by a numerical measure using serial serum creatinine, allow for an AKI prediction algorithm that is reproducible on a large scale. Machine learning with recurrent neural network–based techniques could improve the accuracy of analytics over traditional biostatistics [12]. These could be enhanced by capturing the relative feature importance temporal to AKI; that is, certain clinical covariates or trends (ie, features) would factor with increasing (or decreasing) importance in the time leading up to the onset of AKI. In this study, we would apply a novel machine learning technique that analyzes patient-related features in the form of routine hematology and biochemistry and their interaction with time to accurately predict AKI in hospitals by a lead time of 48 hours.

Methods

Dataset

The data source was our institution's EHR in 2012, which recorded clinical and laboratory data from 68,832 case admissions in that year. Our institution is a 1200-bed academic hospital that provides complex tertiary care services including cardiothoracic surgery, transplantation, and cancer management. The Institutional Human Research Ethics Committee approved the study (NUHS-DSRB 2018/00169) and waived the need for informed consent given the use of deidentified data for analytics with secured institutional governance.

Study Design and Participants

We performed an observational longitudinal study of the prospectively acquired EHR data from hospitalized patients in 2012. The exclusion criteria were (1) patients discharged within 48 hours of admission; (2) patients with community-acquired AKI, as inferred from onset of AKI within 48 hours of hospitalization [13]; (3) patients with stage 5 chronic kidney disease by Kidney Disease: Improving Global Outcomes (KDIGO) criteria, both dialysis, and nondialysis [14], inferred from diagnosis codes (Systematized Nomenclature of Medicine–Clinical Terms) for "end-stage kidney/renal disease," an admission estimated glomerular filtration rate (eGFR) of less than 15 mL/min/1.73 m^2 by Chronic Kidney Disease Epidemiology Collaboration equation [15], or procedural codes for peritoneal dialysis catheter insertion, arteriovenous access creation, or fistuloplasty; (4) patients with procedural codes for "**dialysis," "**filtration," or "**diafiltration" previously who failed to recover kidney function to a current admission eGFR of at least 30 mL/min/1.73 m^2; or (5) patients with no available laboratory results for analytics.

Definition of AKI

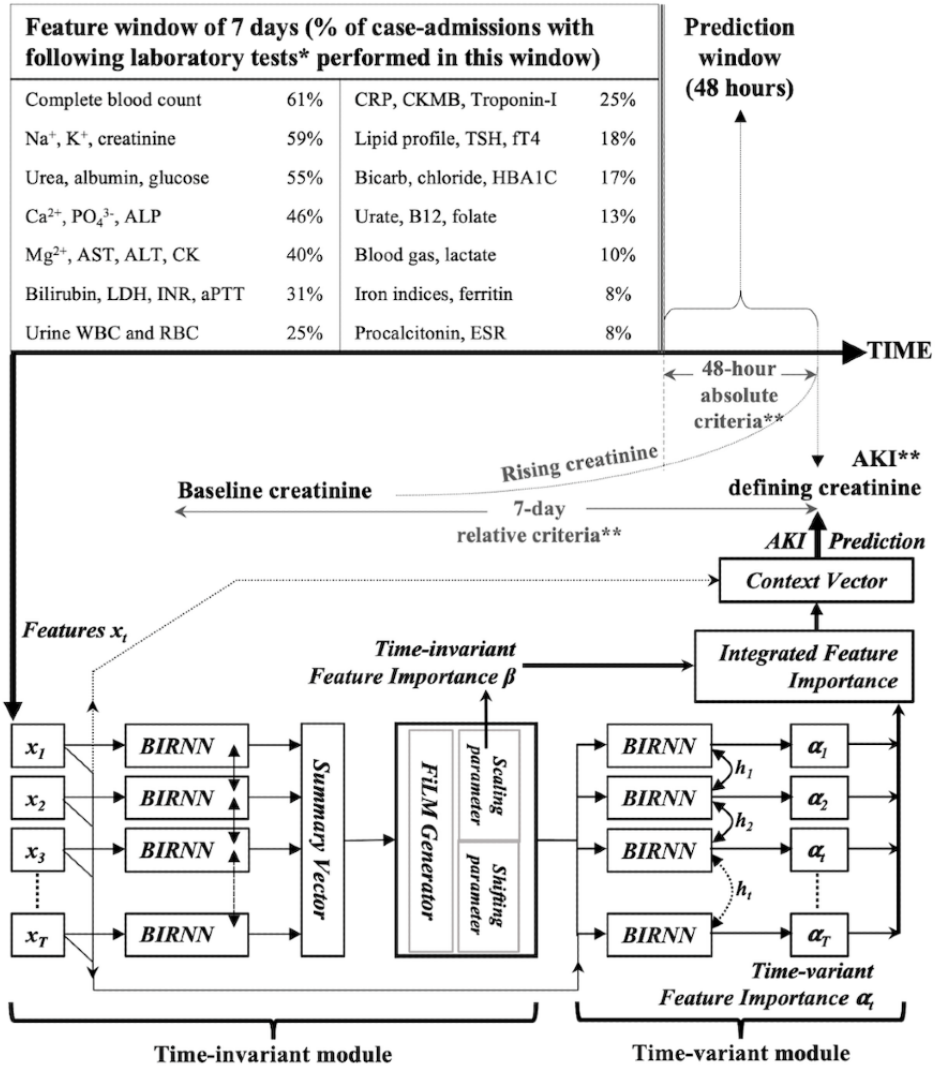
The binary event measure was AKI, as defined by the KDIGO 2012 criteria using serial serum creatinine levels during the index hospitalization [16]. These included the relative criterion of at least 1.5 times an increase in serum creatinine level within a 7-day window; the absolute criterion was an increase in serum creatinine of greater than 26.5 μmol/L (0.3 mg/dL) within 48 hours. The reference serum creatinine within the corresponding 7-day or 48-hour window for either criterion was taken as the baseline creatinine. The AKI-defining creatinine level and the extent of elevation over baseline were used to grade the initial

KDIGO AKI staging severity. Creatinine was measured using the ADVIA 2400 (Siemens AG) enzymatic method traceable to isotope dilution mass spectrometry standard. We did not apply the oliguria criterion for AKI.

Features Used for Analytics

Features (or covariates) were sourced from time-series laboratory results. The data source was our institution's EHR, Computerized Patient Support System version 2 (Integrated Health Information System Pte Ltd). The results were integrated from comma-separated value files using common masked identifiers and ported onto our institution's artificial intelligence discovery platform, an EHR analytic module. Data with the date and time stamps were selected as features to predict the event. These included all serial hematology, serum biochemistry, and urinary investigations (Figure 1). We did not include disease diagnosis codes or medication records.

Figure 1. Prediction logic and features included in analytics. *: Serum biochemistry or hematology unless otherwise stated (eg, urine WBC and RBC); **: AKI defined by KDIGO criteria; x: features entered in model; t: time windows; β: time-invariant feature importance of which influence is shared across time windows; alpha-t: time-variant feature importance; ht: time-variant hidden representation; WBC: white blood cell; RBC: red blood cell; AKI: acute kidney injury; KDIGO: Kidney Disease: Improving Global Outcomes; BIRNN: bidirectional recurrent neural network; FiLM: feature-wise linear modulation.



Analytics

Patient profile was compared between unique patients who developed AKI and those who did not. Parametric variables were reported as mean and standard deviation and compared using Student *t* tests; nonparametric variables were reported as median and IQR and compared using Wilcoxon rank-sum tests. Categorical variables were reported as frequency and percentage and compared using chi-square or Fisher exact tests where appropriate. A 2-tailed *P* value of <.05 was taken as the measure of statistical significance.

We sectioned the dataset by date and time for predictive analytics. Every case admission was taken as one sample. The first AKI episode that occurred in corresponding case admissions was analyzed. The AKI-defining creatinine served as the reference time point; the immediately preceding 48 hours was made the prediction window, and the feature window included the time up to 7 days before the prediction window (Figure 1). For case admissions with no AKI episodes, the corresponding

reference time point would be the final serum creatinine level and likewise preceded by a 48-hour prediction window and a further 7-day feature window. Features performed within the feature window were used to predict AKI, a bivariate event, by a lead time of 48 hours. The feature window was further sectioned into daily serial time intervals for time-series modeling, temporal to the event. For each time interval, we averaged the values of the same feature, followed by normalization of the corresponding result x to generate a normalized x^1 as the input for analytics, where $x^1 = [x - \text{minimum}(x)] / [\text{maximum}(x) - \text{minimum}(x)]$.

We proposed a novel time-invariant and time-variant (TITV) model to facilitate more accurate and interpretable analytics in AKI prediction based on the collaboration of 3 modules [17] (Figure 1). In the time-invariant module, an abstract representation was calculated with the data in the entire feature window, denoting each feature's importance shared across time (ie, time-invariant feature importance). This time-invariant feature importance guided the modulation of input in the next module, the time-variant module. In this second module, we applied a bidirectional recurrent neural network to process sequential data and capture the dynamic behavior both forward and backward in time temporal to the event, as guided by the computed time-invariant feature importance from the time-invariant module. Additionally, we differentiated the influence of features across time windows leading to the event by applying the self-attention mechanism on top of the output of the bidirectional recurrent neural network; the output after the self-attention mechanism represents each feature's importance in the corresponding time window (ie, time-variant feature importance in this time-variant module). Finally, in the prediction module, both the time-invariant and the time-variant feature importance were aggregated to calculate the final prediction (ie, risk of AKI). Meanwhile, the influence of each feature (in each time window) on the final prediction was also derived from the TITV model.

We performed a random shuffling of the entire cohort and arbitrarily partitioned the samples into 80% training, 10% validation, and 10% testing datasets. In the training process, we selected the hyperparameters that achieved the best performance in the validation dataset and applied them to the testing dataset for reporting of the experimental results [18-20]. We examined the reporting performance using the area under the receiver operating characteristic curve (AUC), as well as the respective sensitivity (recall) and positive predictive values (precision) that corresponded with the varying model prediction thresholds for AKI. Precision represents the proportion of predicted cases that truly had AKI; recall represents the proportion of actual AKI cases successfully identified by the prediction model. The AKI prediction threshold that provided the most optimal statistical balance between precision and recall was inferred by the highest computed F1 score. A high model recall gives rise, however, to more false positives (ie, poorer precision), and these permutations were further examined to demonstrate their clinical relevance to AKI diagnostics. These results were compared with the corresponding performance using traditional logistic regression and baseline recurrent neural network models. We applied zero imputation for missing data. Analysis was performed using Python (version 3.8.2, open source for Mac OSX).

Results

Patient Profile

We studied 20,732 case admissions in 16,288 unique patients, of which 1971 patients were younger than age 18 years (Figure 2). The mean age of the final cohort was 53 (SD 25) years, and 52.2 (8510/16,288) were males; 28.9% (4701/16,288) had diabetes, 35.0% (5699/16,288) had hypertension, 11.7% (1898/16,288) had ischemic heart disease, and 11.7% (1899/16,288) had either solid organ or hematological malignancy. Near half (7214/16,288, 44.3%) of patients had a baseline eGFR<90 mL/min/1.73 m^2. More patients with AKI (258/869, 29.7%) had a baseline eGFR<60 mL/min/1.72 m^2 compared to those without AKI (2738/15,419, 17.8%; P<.001; Table 1).

Figure 2. Study flow diagram. AKI: acute kidney injury; CKD: chronic kidney disease; CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration equation; eGFR: estimated glomerular filtration rate; ESKD: end-stage kidney disease; RRT: renal replacement therapy.

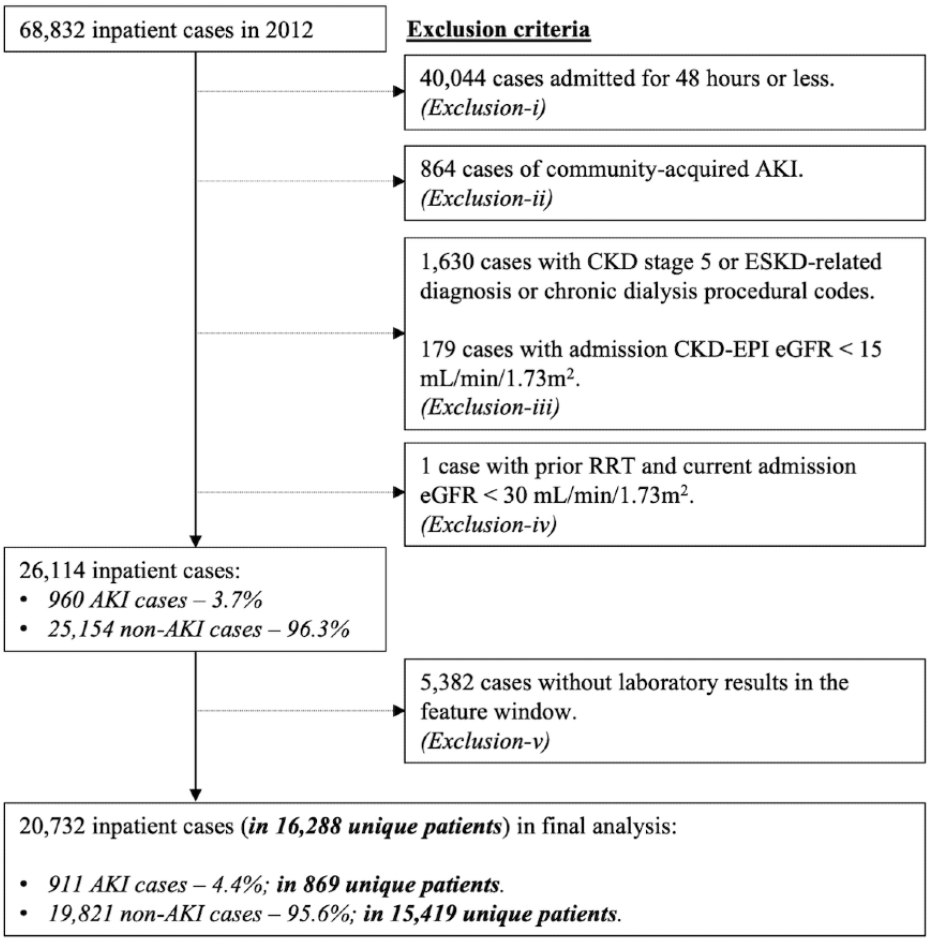


Table 1. Study profile and bivariate comparison between acute kidney injury and non–acute kidney injury patients.

Variables	Entire cohort (n=16,288)	AKI[a] (n=869)	Non-AKI (n=15,419)	*P* value
Age, mean (SD), years	53 (25)	62 (22)	53 (26)	<.001
Male gender, n (%)	8510 (52.2)	480 (55.2)	8030 (52.1)	.08
Comorbidities, n (%)				
Diabetes	4701 (28.9)	371 (42.7)	4330 (28.1)	<.001
Hypertension	5699 (35.0)	498 (57.3)	5201 (33.7)	<.001
Ischemic heart disease	1898 (11.7)	262 (30.1)	1636 (10.6)	<.001
Heart failure	1138 (7.0)	190 (21.9)	948 (6.1)	<.001
Cerebrovascular disease	757 (4.6)	73 (8.4)	684 (4.4)	<.001
Chronic liver disease	269 (1.7)	44 (5.1)	225 (1.5)	<.001
Solid organ malignancy	1556 (9.6)	140 (16.1)	1416 (9.2)	<.001
Hematological malignancy	343 (2.1)	52 (6.0)	291 (1.9)	<.001
Baseline kidney function				
Creatinine, μmol/L, median (IQR)	71 (54-92)	69 (46-108)	71 (55-91)	.05
eGFR[b], mL/min/1.73 m^2, median (IQR)	91 (67-109)	86 (55-110)	91 (68-109)	<.001
eGFR 90 or above mL/min/1.73 m^2, n (%)	7628 (46.8)	400 (46.0)	7228 (46.9)	.65
eGFR 60 to <90 mL/min/1.73 m^2, n (%)	4218 (25.9)	211 (24.3)	4007 (26.0)	.28
eGFR 45 to <60 mL/min/1.73 m^2, n (%)	1328 (8.2)	101 (11.6)	1227 (8.0)	<.001
eGFR 30 to <45 mL/min/1.73 m^2, n (%)	950 (5.8)	90 (10.4)	860 (5.6)	<.001
eGFR <30 mL/min/1.73 m^2, n (%)	718 (4.4)	67 (7.7)	651[c] (4.2)	<.001
AKI-related variables				
AKI-defining creatinine, μmol/L, median (IQR)	—[d]	122 (80-169)	—	—
Relative criterion (vs absolute), n (%)	—	651 (74.9)	—	—
AKI onset days from admission, median (IQR)	—	6 (3-10)	—	—
Serum biochemistry at AKI detection, median (IQR)				
Sodium, mmol/L	—	138 (135-142)	—	—
Potassium, mmol/L	—	4.1 (3.7-4.6)	—	—
Urea, mmol/L	—	11 (7-15)	—	—
Bicarbonate, mmol/L	—	24 (19-27)	—	—
Phosphate, mmol/L	—	1.23 (0.95-1.54)	—	—
Calcium, mmol/L	—	2.03 (1.89-2.17)	—	—
Chloride, mmol/L	—	105 (101-109)	—	—
Uric acid, μmol/L	—	384 (266-527)	—	—
Initial KDIGO[e] AKI staging, n (%)				
Stage 1	—	701 (80.7)	—	—
Stage 2	—	125 (14.4)	—	—
Stage 3	—	43 (4.9)	—	—
Total cumulative hospital days, median (IQR)	5 (3-10)	23 (13-44)	5 (3-9)	<.001
Hospital days per admission, median (IQR)	5 (3-8)	14 (8-26)	5 (3-7)	<.001

[a]AKI: acute kidney injury.

[b]eGFR: estimated glomerular filtration rate by Chronic Kidney Disease Epidemiology Collaboration equation.

[c]A total of 1446 non-AKI patients had missing baseline eGFR.

[d]Not applicable.

[e]KDIGO: Kidney Disease: Improving Global Outcomes.

Evaluation Outcomes

AKI developed during 4.4% (911/20,732) of case admissions in 869 unique patients at a median of 6 (IQR 3-10) days from admission; 74.9% (651/869) of AKI patients were diagnosed based on KDIGO relative criterion, and 80.7% (701/869) were of initial KDIGO stage 1 in severity. Patients who developed AKI were older with more comorbidities including diabetes, hypertension, cardiovascular diseases, chronic kidney disease, chronic liver disease, and cancers compared with non-AKI patients (all $P<.001$). The median hospital days per admission and cumulatively in 2012 in AKI patients versus those without were 14 (IQR 8-26) days versus 5 (IQR 3-7) days, and 23 (IQR 13-44) days versus 5 (IQR 3-9) days, respectively (all $P<.001$; Table 1).

Analytics for AKI Prediction in the Hospital

The 7-day feature window was divided into daily time windows, giving a total of 7 time windows and 709 features in the analysis. Figure 1 shows the laboratory variables included in the feature window in order of their corresponding test prevalence by categories. Complete blood count was the most common investigation, performed in 61.3% (12,709/20,732) of all case admissions in the analysis; this was followed by serum electrolytes, urea, and creatinine at 46% to 59%, and liver function markers at 30% to 41%. In comparison, acid-base parameters and serum lactate contributed less (2146/20,732, 10.4%) to the analysis.

The cohort was partitioned into the training (16,585 cases), validation (2073 cases), and testing (2074 cases) datasets; AKI rates in the 3 datasets were 4.5%, 3.9%, and 4.3%, respectively. Table 2 summarizes the AUC of respective analytic modules in the final testing dataset as well as the precision and recall corresponding with the AKI prediction threshold with the highest F1 score. The AUC for AKI prediction by the multivariate logistic regression and recurrent neural network/time-series models were 79% and 80%, respectively. The AUC was 81% after we applied the TITV module with comparable precision and recall compared with the former models; these and the highest F1 score were achieved at an AKI prediction threshold between 15% and 20%. The respective AUCs and corresponding area under precision-recall curves for the training and testing datasets are illustrated in Figure 3.

Table 3 shows the breakdown in our TITV module precision and recall according to the varying probability thresholds for AKI prediction.

A low prediction threshold detected a very high number of predicted AKI cases that scored high in recall but poor in discrimination between true and false positives. Conversely, a high prediction threshold detected a low number of predicted AKI cases but with high precision. A 15% AKI probability threshold implied that 699 cases were predicted to be diagnosed with AKI; 33.3% (233/699) of predicted cases did subsequently develop AKI, while 25.6% (233/911) of eventual AKI cases were successfully predicted. Reducing the probability threshold to 5% led to 3746 predicted AKI cases with much higher false positives but with successful prediction of 60.0% (547/911) of eventual AKI cases. Figure 4 illustrates the confusion matrix plots at AKI prediction thresholds of 5% and 15%. Further details on TITV performance metrics are provided in Table 4.

In addition, our TITV model generated representative interpretation results specific to each AKI case. Figure 5 illustrates the relative feature importance to AKI in 8 case examples, which demonstrated the range of inflammatory, cardiac, drug-specific, or hepatic functional markers in association with AKI, specific to each case. The source codes for our predictive algorithm are available online [21].

Table 2. Acute kidney injury predictive performance in the testing dataset with optimized F1.

Model	Precision[a]	Recall[b]	F1[c]	AUC[d] (95% CI)
Logistic regression	0.274	0.189	0.224	0.789 (0.752-0.827)
RNN[e] (GRU[f])	0.286	0.222	0.250	0.800 (0.764-0.836)
BRNN[g] (BGRU[h])	0.309	0.233	0.266	0.797 (0.761-0.833)
Proposed TITV[i] model	0.397	0.256	0.311	0.814 (0.780-0.848)

[a]Precision: true positive / (all cases predicted at risk of acute kidney injury).

[b]Recall: true positive / (all cases that eventually developed acute kidney injury).

[c]F1 score: 2 × [(recall × precision) / (recall + precision)].

[d]AUC: area under receiver operating characteristic curve.

[e]RNN: recurrent neural network.

[f]GRU: gated recurrent unit.

[g]BRNN: bidirectional recurrent neural network.

[h]BGRU: bidirectional gated recurrent unit.

[i]TITV: time-invariant and time-variant feature importance.

Figure 3. Area under receiver operating characteristic and area under precision-recall curves of training and testing datasets. AUC: area under receiver operating characteristic curve; AUPRC: area under precision-recall curve.

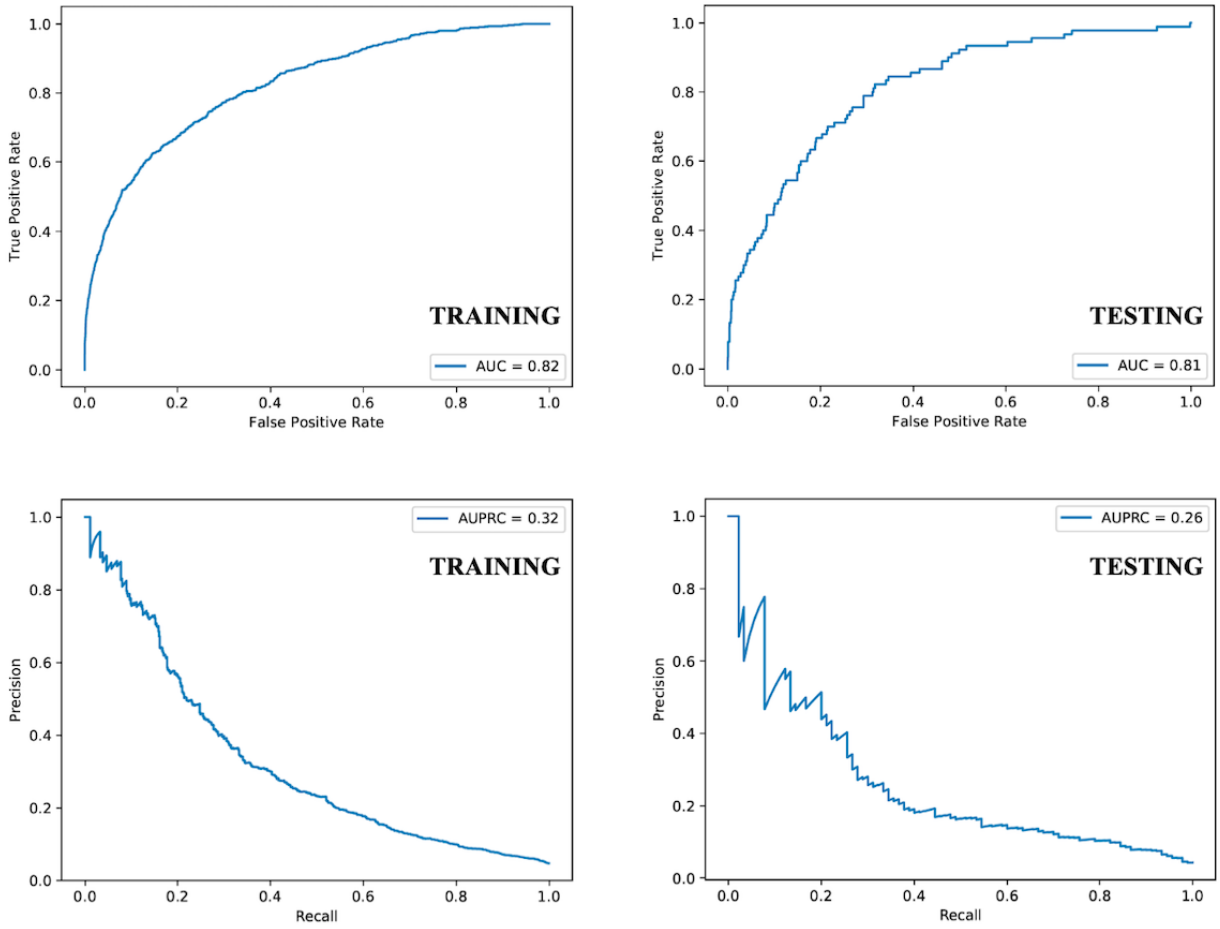


Table 3. Varied acute kidney injury prediction thresholds on time-invariant and time-variant model performance metrics.

Threshold[a] to predict AKI[b] (%)	Precision[c]	Recall[d]	F1[e]	Predicted AKI cases by model, n	True positive AKI cases, n
5	0.146	0.600	0.235	3746	547
10	0.252	0.333	0.287	1204	304
15	0.333	0.256	0.289	699	233
20	0.500	0.200	0.286	364	182
25	0.480	0.133	0.209	253	121
30	0.556	0.111	0.185	182	101

[a]Probability threshold to define predicted AKI versus no risk of AKI (ie, positive/negative class prediction). A low threshold risks over-detection and alert fatigue, which corresponds to poor precision. A high threshold risks missing true AKI cases, which corresponds to poor recall.

[b]AKI: acute kidney injury.

[c]Precision: true positive / (all cases predicted at risk of AKI).

[d]Recall: true positive / (all cases who eventually developed AKI).

[e]F1 score: 2 × [(recall × precision) / (recall + precision)].

Figure 4. Confusion matrix plots with acute kidney injury prediction thresholds at 5% and 15%.

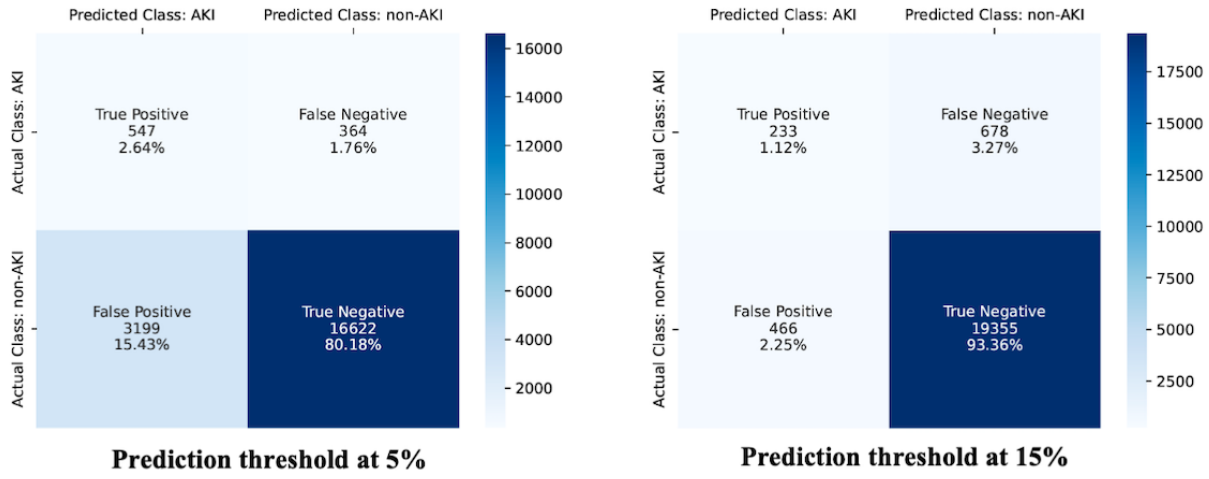


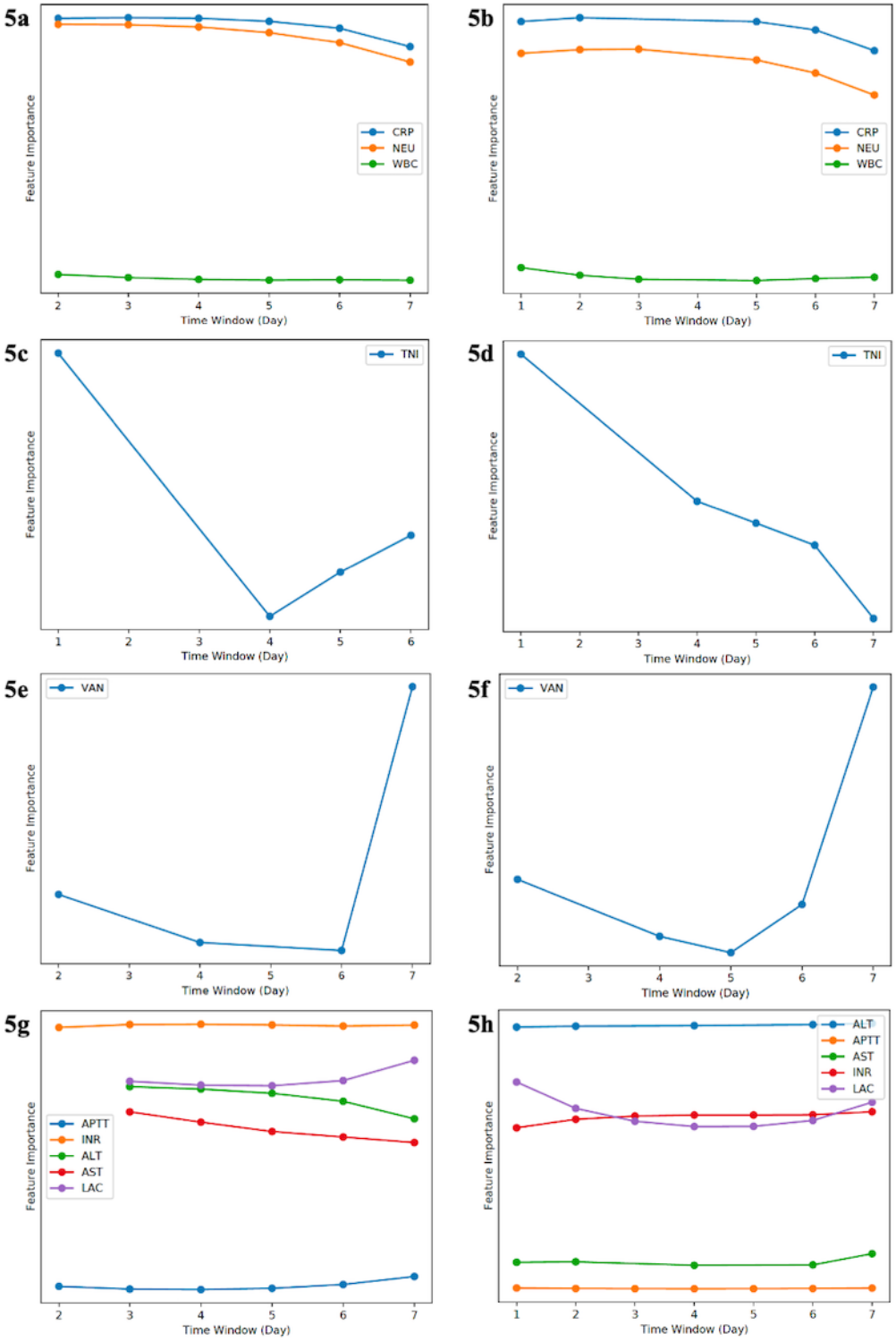
Table 4. Model performance metric with time-invariant and time-variant prediction thresholds at 5% and 15%.

	True AKI[a] cases	No AKI	Subtotal
5% prediction threshold			
TITV[b] predicted (positive)	547	3199	3746
TITV predicted (negative)	364	16,622	16,986
Subtotal	911	19,821	20,732
15% prediction threshold			
TITV predicted (positive)	233	466	699
TITV predicted (negative)	678	19,355	20,033
Subtotal	911	19,821	20,732

[a]AKI: acute kidney injury.

[b]TITV: time-invariant and time-variant module.

Figure 5. Case examples of relative feature importance in acute kidney injury (AKI) prediction. Time-window: refers to feature window of 7 days in AKI prediction; Y-axis: features highly associated with AKI would rank high in relative feature importance; a-b: C-reactive protein, neutrophils featured prominently over days, which suggested infection and inflammation were associated with subsequent AKI; c-d: troponin-I featured prominently initially, which suggested cardiac disease in association with AKI, although its relative importance waned in subsequent days; e-f: vancomycin levels rose in feature importance proximate to AKI, which strongly suggested vancomycin nephrotoxicity; g-h: lactate, liver enzymes, international normalized ratio, and activated partial thromboplastin time featured strongly, which suggested hepatic or multiorgan dysfunction in association with evolving AKI.



Discussion

Principal Findings

We have used structured but heterogeneous biochemical data from 20,732 case admissions in the prediction of hospital-acquired AKI by a 48-hour lead time. We enhanced the recurrent neural network model with a novel analytic module that took into account the temporal interactions in serial laboratory parameters that inferred disease trajectory leading up to AKI [17]. At the optimal statistical operation point as indicated by the highest F1 score (Table 2), our module generated 3 false positives for every 2 true AKI cases, and clinicians would need to act on just 600 predicted AKI alerts of 20,732 case admissions yearly; however, 3 of 4 true AKI cases would be missed. It may be more desirable for our module to successfully predict at least 3 of 5 true AKI cases, but this is counterbalanced by 6 false positives for every 1 true AKI case, and more than 3000 predicted AKI alerts yearly (Table 3). We suggest that our AKI prediction threshold should be low to identify more patients at risk of AKI daily. This narrows the entire hospital cohort to a more manageable patient number for closer monitoring, in whom further assessment could be augmented by AKI biomarkers to reduce false positives [22]. These include urinary clusterin, kidney injury molecule–1, tissue inhibitor of metalloproteinase–2, and insulin-like growth factor binding protein–7, for which levels rise in 12 to 48 hours before a significant rise in serum creatinine [7,23].

Comparison With Prior Work

Our methodology differs from machine learning techniques that used a quasi-random selection of variable prediction points relative to AKI [24]. It resembles models that adopted structured feature and prediction windows relative to AKI that facilitate the deployment of our prediction algorithm in real time [25]. Importantly, we expanded the prediction window to a minimum of 48 hours. Such improved lead time may be necessary for any AKI preventive strategies to make a meaningful change in clinical outcomes. Preemptive interventions may include more detailed patient reviews, timely treatment of infections, precise volume management [26], preferred use of balanced electrolyte over chloride-rich solutions [27], admission to high-dependency or intensive care unit for detailed monitoring, and reduction in or cessation of nephrotoxic medications [28]. These measures, when implemented in a timely fashion and supported by a responsive EHR platform for AKI alerts, may reduce the hospital days and AKI duration in affected patients [29,30].

The performance of any analytic module depends strongly on the appropriate feature selection. Our model was built from objective laboratory test results that would be similar in data structure across institutions [31]. Our algorithm used routinely performed hematology and biochemistry without disease diagnosis codes; these included complete blood count, common electrolytes, acid-base parameters, and liver and cardiac enzymes, and these remain relevant for current AKI prediction even with the changing health care landscape. As our analysis was limited to available investigations performed before a mandatory 48-hour prediction window, the laboratory indices analyzed in the feature window might not be comprehensive. This could compromise the model performance, and the prediction should otherwise improve with features performed at higher frequency and more proximate to AKI [25,32]. Despite this, we demonstrated an AUC that exceeded 80% for AKI prediction in our testing dataset. Certain indices like blood gas, serum lactate, cardiac enzymes, and drug levels should increase in frequency and importance toward the onset of AKI, since AKI serves as a marker of clinical deterioration from nosocomial infections, decompensated cardiovascular diseases, major surgery, or nephrotoxicity [33,34]. Varying significance of these time-sensitive features in association with evolving AKI may be seen among subsets of patients with sepsis, cardiac failure, or cardiac surgery [35-37]. Our TITV module can provide patient-level interpretation of the feature importance, as suggested by our representative interpretation results in unique AKI case examples (Figure 5). These could provide insightful patient-specific trends to aid the evaluation of AKI etiology [17].

Strengths and Limitations

Our study has several strengths but is not without limitations. We have studied a large and diverse population with a comprehensive range of medical and surgical conditions not confined to critical care, which improves the generalizability of our analytic module to hospital practice. We excluded patients with more advanced chronic kidney disease, and our 4% incidence of AKI in the hospital was lower than the 8% reported in prior studies that used similar EHR methods [33,38]. The lack of precise urine output in ward patients could reduce the model accuracy, but oliguria often develops in 24 hours proximate to AKI and may not fulfill our requirement for a 48-hour prediction window. We have normalized the variables for standardized comparison across different tests. Our novel TITV module provided fine-grained interpretability of the prediction results and achieved accurate prediction simultaneously; this facilitates high-quality health care analytics. Being single center in nature, our AKI prediction module needs to be applied and validated in external health care systems to demonstrate reproducibility. The prediction algorithm could be ported to run on platforms that use similar EHR data architecture, but this naturally limits its deployment to institutions with available technology. Nevertheless, our model could be applied for rolling AKI predictions daily if coupled with a real-time feed of laboratory data. While forward application of the algorithm would naturally encounter model degradation due to concept drift, novel techniques could achieve concept drift detection, understanding, and further adaption from contemporaneous data [39,40]. Furthermore, our algorithm was based on laboratory test results less subjected to case-mix shift over time as compared with disease diagnoses or medication records [41]. We had used zero imputation for missing data, unlike the previously described method of imputing preexisting values in time or median value [38]; zero imputation has been widely adopted in machine learning techniques and has achieved state-of-the-art performance in analytics [42,43]. Finally, the subcohort with “false-positive AKI” might be analogous to that of patients with subclinical AKI that may also be associated with adverse long-term outcomes; these were not explored in our study.

Conclusions

We have presented a feasible and enhanced EHR analytic module that captures time-sensitive interactions in laboratory investigations and predicts hospital-acquired AKI by a 48-hour lead time. The AKI prediction threshold could be varied to allow the clinically relevant balance in model precision, recall, and predicted AKI numbers that are compatible with patient service load in health care institutions. With a compromised precision in favor of the better recall, our model serves to risk stratify ward patients for detailed clinical or biomarker assessment for true AKI risk. Its real-time deployment is expected to greatly facilitate our upstream efforts to prevent AKI or its complications in hospitalized patients.

Acknowledgments

This study was supported by a grant from Artificial Intelligence Singapore (100 Experiments Grant), a program supported by the National Research Foundation and hosted by the National University of Singapore. The granting body was not involved in the review or approval of the manuscript for publication. The publication was supported by the Data and Technology Office, National University Health System, Singapore. We thank our institution's EHR data managers, Mr Khong-Hai Wang and Mr Mohammad Shaheryar Furqan, and administrator, Ms Mar Mar Win, for their invaluable support.

Authors' Contributions

HRC, AV, KYN, and HYT conceived the study concept. HRC, KZ, KYN, KA, and BCO planned the methodology and obtained the ethics approval. HRC, KZ, KYN, KA, and BCO retrieved the EHRs. HRC, KZ, AV, HKY, LL, HYT, AM, GM, and SLL preprocessed and curated the data. KZ and BCO applied the machine learning techniques and analyzed the data. HRC, KZ, AV, HKY, LL, HYT, AM, GM, and SLL performed the interim review of analytics. HRC, KZ, and BCO wrote the manuscript. All authors were involved in revising the manuscript and take responsibility for the data presented.

Conflicts of Interest

None declared.

References

1. Fabbian F, Savriè C, De Giorgi A, Cappadona R, Di Simone E, Boari B, et al. Acute kidney injury and in-hospital mortality: a retrospective analysis of a nationwide administrative database of elderly subjects in Italy. J Clin Med 2019 Sep 02;8(9):1 [FREE Full text] [doi: 10.3390/jcm8091371] [Medline: 31480750]
2. Christiansen CF, Johansen MB, Langeberg WJ, Fryzek JP, Sørensen HT. Incidence of acute kidney injury in cancer patients: a Danish population-based cohort study. Eur J Intern Med 2011 Aug;22(4):399-406. [doi: 10.1016/j.ejim.2011.05.005] [Medline: 21767759]
3. Pavkov ME, Harding JL, Burrows NR. Trends in hospitalizations for acute kidney injury—United States, 2000-2014. MMWR Morb Mortal Wkly Rep 2018 Mar 16;67(10):289-293 [FREE Full text] [doi: 10.15585/mmwr.mm6710a2] [Medline: 29543788]
4. Low S, Vathsala A, Murali TM, Pang L, MacLaren G, Ng W, et al. Electronic health records accurately predict renal replacement therapy in acute kidney injury. BMC Nephrol 2019 Jan 31;20(1):32 [FREE Full text] [doi: 10.1186/s12882-019-1206-4] [Medline: 30704418]
5. Bagshaw SM, Uchino S, Bellomo R, Morimatsu H, Morgera S, Schetz M, Beginning and Ending Supportive Therapy for the Kidney (BEST Kidney) Investigators. Septic acute kidney injury in critically ill patients: clinical characteristics and outcomes. Clin J Am Soc Nephrol 2007 May;2(3):431-439 [FREE Full text] [doi: 10.2215/CJN.03681106] [Medline: 17699448]
6. Ong L, Tambyah PA, Lum LH, Low Z, Cheng I, Murali TM, et al. Aminoglycoside-associated acute kidney injury in elderly patients with and without shock. J Antimicrob Chemother 2016 Nov;71(11):3250-3257. [doi: 10.1093/jac/dkw296] [Medline: 27494924]
7. Da Y, Akalya K, Murali T, Vathsala A, Tan C, Low S, et al. Serial quantification of urinary protein biomarkers to predict drug-induced acute kidney injury. Curr Drug Metab 2019;20(8):656-664. [doi: 10.2174/1389200220666190711114504] [Medline: 31296157]
8. Chong E, Poh KK, Liang S, Soon CY, Tan H. Comparison of risks and clinical predictors of contrast-induced nephropathy in patients undergoing emergency versus nonemergency percutaneous coronary interventions. J Interv Cardiol 2010 Oct;23(5):451-459 [FREE Full text] [doi: 10.1111/j.1540-8183.2010.00581.x] [Medline: 20796168]
9. Chong E, Poh K, Lu Q, Zhang JJ, Tan N, Hou XM, et al. Comparison of combination therapy of high-dose oral N-acetylcysteine and intravenous sodium bicarbonate hydration with individual therapies in the reduction of Contrast-induced Nephropathy during Cardiac Catheterisation and Percutaneous Coronary Intervention (CONTRAST): a multi-centre, randomised, controlled trial. Int J Cardiol 2015 Dec 15;201:237-242. [doi: 10.1016/j.ijcard.2015.07.108] [Medline: 26301645]

10. Chua H, Wong W, Ong VH, Agrawal D, Vathsala A, Tay H, et al. Extended mortality and chronic kidney disease after septic acute kidney injury. J Intensive Care Med 2020 Jun;35(6):527-535. [doi: 10.1177/0885066618764617] [Medline: 29552953]
11. Hong W, Haroon S, Lau T, Hui S, Ma V, Goh J, et al. Transitional care program to facilitate recovery following severe acute kidney injury. J Nephrol 2019 Aug;32(4):605-613. [doi: 10.1007/s40620-019-00616-z] [Medline: 31090023]
12. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: predicting clinical events via recurrent neural networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
13. Wonnacott A, Meran S, Amphlett B, Talabani B, Phillips A. Epidemiology and outcomes in community-acquired versus hospital-acquired AKI. Clin J Am Soc Nephrol 2014 Jun 06;9(6):1007-1014 [FREE Full text] [doi: 10.2215/CJN.07920713] [Medline: 24677557]
14. Willis K, Cheung M, Slifer S. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Disease: Improving Global Outcomes. 2013. URL: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf [accessed 2021-12-02]
15. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med 2009 May 05;150(9):604-612 [FREE Full text] [doi: 10.7326/0003-4819-150-9-200905050-00006] [Medline: 19414839]
16. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. Nephron Clin Pract 2012;120(4):c179-c184 [FREE Full text] [doi: 10.1159/000339789] [Medline: 22890468]
17. Zheng K, Cai S, Chua H, Wang W, Ngiam K, Ooi B. Tracer: a framework for facilitating accurate and interpretable analytics for high stakes applications. Proc 2020 ACM SIGMOD Int Conf Manag Data 2020:1747-1763. [doi: 10.1145/3318464.3389720]
18. Choi E, Bahadori M, Kulas J, Schuetz A, Stewart W, Sun J. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. Proc 30th Int Conf Neural Inf Proc Syst 2016:3512-3520. [doi: 10.5555/3157382.3157490]
19. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. ArXiv. Preprint posted online on June 19, 2017 2017:1903-1911 [FREE Full text] [doi: 10.1145/3097983.3098088]
20. Bai T, Egleston BL, Zhang S, Vucetic S. Interpretable representation learning for healthcare via capturing disease progression through time. KDD 2018 Aug:43-51 [FREE Full text] [doi: 10.1145/3219819.3219904] [Medline: 31037221]
21. kaiping/TRACER. URL: https://github.com/kaiping/TRACER [accessed 2021-12-02]
22. Liu KD, Goldstein SL, Vijayan A, Parikh CR, Kashani K, Okusa MD, AKI!Now Initiative of the American Society of Nephrology, et al. AKI!Now initiative: recommendations for awareness, recognition, and management of AKI. Clin J Am Soc Nephrol 2020 Dec 07;15(12):1838-1847. [doi: 10.2215/CJN.15611219] [Medline: 32317329]
23. Kashani K, Al-Khafaji A, Ardiles T, Artigas A, Bagshaw SM, Bell M, et al. Discovery and validation of cell cycle arrest biomarkers in human acute kidney injury. Crit Care 2013 Mar 06;17(1):R25 [FREE Full text] [doi: 10.1186/cc12503] [Medline: 23388612]
24. Rank N, Pfahringer B, Kempfert J, Stamm C, Kühne T, Schoenrath F, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. NPJ Digit Med 2020;3:139 [FREE Full text] [doi: 10.1038/s41746-020-00346-8] [Medline: 33134556]
25. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019 Aug;572(7767):116-119. [doi: 10.1038/s41586-019-1390-1] [Medline: 31367026]
26. Meersch M, Schmidt C, Hoffmeier A, Van Aken H, Wempe C, Gerss J, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. Intensive Care Med 2017 Nov;43(11):1551-1561 [FREE Full text] [doi: 10.1007/s00134-016-4670-3] [Medline: 28110412]
27. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, SALT-ED Investigators. Balanced crystalloids versus saline in noncritically ill adults. N Engl J Med 2018 Mar 01;378(9):819-828 [FREE Full text] [doi: 10.1056/NEJMoa1711586] [Medline: 29485926]
28. Goldstein SL, Mottes T, Simpson K, Barclay C, Muething S, Haslam DB, et al. A sustained quality improvement program reduces nephrotoxic medication-associated acute kidney injury. Kidney Int 2016 Dec;90(1):212-221. [doi: 10.1016/j.kint.2016.03.031] [Medline: 27217196]
29. Kolhe NV, Reilly T, Leung J, Fluck RJ, Swinscoe KE, Selby NM, et al. A simple care bundle for use in acute kidney injury: a propensity score-matched cohort study. Nephrol Dial Transplant 2016 Nov;31(11):1846-1854. [doi: 10.1093/ndt/gfw087] [Medline: 27190331]
30. Selby NM, Casula A, Lamming L, Stoves J, Samarasinghe Y, Lewington AJ, et al. An organizational-level program of intervention for AKI: a pragmatic stepped wedge cluster randomized trial. J Am Soc Nephrol 2019 Mar;30(3):505-515 [FREE Full text] [doi: 10.1681/ASN.2018090886] [Medline: 31058607]
31. Jones RG, Johnson OA, Batstone G. Informatics and the clinical laboratory. Clin Biochem Rev 2014 Aug;35(3):177-192 [FREE Full text] [Medline: 25336763]

32. Cheng P, Waitman LR, Hu Y, Liu M. Predicting inpatient acute kidney injury over different time horizons: how early and accurate? AMIA Annu Symp Proc 2017;2017:565-574 [FREE Full text] [Medline: 29854121]
33. Jannot A, Burgun A, Thervet E, Pallet N. The diagnosis-wide landscape of hospital-acquired AKI. Clin J Am Soc Nephrol 2017 Jun 07;12(6):874-884 [FREE Full text] [doi: 10.2215/CJN.10981016] [Medline: 28495862]
34. Sykes L, Sinha S, Hegarty J, Flanagan E, Doyle L, Hoolickin C, et al. Reducing acute kidney injury incidence and progression in a large teaching hospital. BMJ Open Qual 2018;7(4):e000308 [FREE Full text] [doi: 10.1136/bmjoq-2017-000308] [Medline: 30555931]
35. Plataki M, Kashani K, Cabello-Garza J, Maldonado F, Kashyap R, Kor DJ, et al. Predictors of acute kidney injury in septic shock patients: an observational cohort study. Clin J Am Soc Nephrol 2011 Jul;6(7):1744-1751 [FREE Full text] [doi: 10.2215/CJN.05480610] [Medline: 21734090]
36. Holgado JL, Lopez C, Fernandez A, Sauri I, Uso R, Trillo JL, et al. Acute kidney injury in heart failure: a population study. ESC Heart Fail 2020 Apr;7(2):415-422 [FREE Full text] [doi: 10.1002/ehf2.12595] [Medline: 32059081]
37. Ng RRG, Chew STH, Liu W, Shen L, Ti LK. Identification of modifiable risk factors for acute kidney injury after coronary artery bypass graft surgery in an Asian population. J Thorac Cardiovasc Surg 2014 Apr;147(4):1356-1361 [FREE Full text] [doi: 10.1016/j.jtcvs.2013.09.040] [Medline: 24183907]
38. Koyner JL, Adhikari R, Edelson DP, Churpek MM. Development of a multicenter ward-based AKI prediction model. Clin J Am Soc Nephrol 2016 Nov 07;11(11):1935-1943 [FREE Full text] [doi: 10.2215/CJN.00280116] [Medline: 27633727]
39. Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: a review. IEEE Trans Knowl Data Eng 2019;31(12):2346-2363. [doi: 10.1109/tkde.2018.2876857]
40. Gama J, Žliobaitė I, Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation. ACM Comput Surv 2014 Apr;46(4):1-37. [doi: 10.1145/2523813]
41. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. J Am Med Inform Assoc 2017 Nov 01;24(6):1052-1061 [FREE Full text] [doi: 10.1093/jamia/ocx030] [Medline: 28379439]
42. Zheng K, Gao J, Ngiam K, Ooi B, Yip W. Resolving the bias in electronic medical records. Proc 23rd ACM SIGKDD Int Con Knowl Discov Data Min 2017:2171-2180. [doi: 10.1145/3097983.3098149]
43. Zheng Y, Tang B, Ding W, Zhou H. A neural autoregressive approach to collaborative filtering. Proc 33rd Int Conf Int Conf Mach Learn 2016:764-773. [doi: 10.5555/3045390.3045472]

Abbreviations

AKI: acute kidney injury
AUC: area under the receiver operating characteristic curve
eGFR: estimated glomerular filtration rate
EHR: electronic health record
KDIGO: Kidney Disease: Improving Global Outcomes
TITV: time-invariant and time-variant

Edited by R Kukafka; submitted 30.05.21; peer-reviewed by Z Ren, M Liu, E Sükei; comments to author 02.08.21; revised version received 26.09.21; accepted 29.10.21; published 24.12.21.

Please cite as:
Chua HR, Zheng K, Vathsala A, Ngiam KY, Yap HK, Lu L, Tiong HY, Mukhopadhyay A, MacLaren G, Lim SL, Akalya K, Ooi BC Health Care Analytics With Time-Invariant and Time-Variant Feature Importance to Predict Hospital-Acquired Acute Kidney Injury: Observational Longitudinal Study
J Med Internet Res 2021;23(12):e30805
URL: https://www.jmir.org/2021/12/e30805
doi:10.2196/30805
PMID:34951595

Original Paper

Prescribing Smartphone Apps for Physical Activity Promotion in Primary Care: Modeling Study of Health Gain and Cost Savings

Leah Grout[1], BSc, MPH, PhD; Kendra Telfer[1], BSocSci (Hons); Nick Wilson[1], MPH; Christine Cleghorn[1], BSc, MSc, PhD; Anja Mizdrak[1], BA, DPhil

Burden of Disease Epidemiology, Equity and Cost-Effectiveness Program, University of Otago Wellington, Wellington, New Zealand

Corresponding Author:
Anja Mizdrak, BA, DPhil
Burden of Disease Epidemiology, Equity and Cost-Effectiveness Program
University of Otago Wellington
23A Mein Street
Newtown
Wellington, 6242
New Zealand
Phone: 64 49186192
Email: anja.mizdrak@otago.ac.nz

Abstract

Background: Inadequate physical activity is a substantial cause of health loss worldwide, and this loss is attributable to diseases such as coronary heart disease, diabetes, stroke, and certain forms of cancer.

Objective: This study aims to assess the potential impact of the prescription of smartphone apps in primary care settings on physical activity levels, health gains (in quality-adjusted life years [QALYs]), and health system costs in New Zealand (NZ).

Methods: A proportional multistate lifetable model was used to estimate the change in physical activity levels and predict the resultant health gains in QALYs and health system costs over the remaining life span of the NZ population alive in 2011 at a 3% discount rate.

Results: The modeled intervention resulted in an estimated 430 QALYs gained (95% uncertainty interval 320-550), with net cost savings of 2011 NZ $2.2 million (2011 US $1.5 million) over the remaining life span of the 2011 NZ population. On a per capita basis, QALY gains were generally larger in women than in men and larger in Māori than in non-Māori. The health impact and cost-effectiveness of the intervention were highly sensitive to assumptions on intervention uptake and decay. For example, the scenario analysis with the largest benefits, which assumed a 5-year maintenance of additional physical activity levels, delivered 1750 QALYs and 2011 NZ $22.5 million (2011 US $15.1 million) in cost savings.

Conclusions: The prescription of smartphone apps for promoting physical activity in primary care settings is likely to generate modest health gains and cost savings at the population level in this high-income country. Such gains may increase with ongoing improvements in app design and increased health worker promotion of the apps to patients.

(J Med Internet Res 2021;23(12):e31702) doi:10.2196/31702

KEYWORDS

physical activity; smartphone apps; mobile health; mHealth; modeling; primary care; mobile phone

Introduction

Inadequate physical activity is a risk factor for coronary heart disease (CHD), diabetes, stroke, and certain forms of cancer [1,2]. The World Health Organization recommends that adults aged 18 to 64 years should complete at least 150 minutes of moderate-intensity aerobic physical activity, at least 75 minutes of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic physical activity each week [3]. Approximately 25% of adults do not meet the recommended level of physical activity worldwide, and it has been estimated that as many as 5 million deaths could be averted each year if the global population were more active [3].

In New Zealand (NZ), >40% of adults are estimated to be insufficiently physically active [4]. CHD, stroke, and diabetes are among the leading causes of health loss in NZ [5], and noncommunicable diseases contribute to marked health

inequalities, with Māori (Indigenous population), Pasifika, and low-income New Zealanders at higher risk for important health conditions [5,6]. Strategies to increase physical activity at the population level are needed to help address this public health concern and reduce health inequalities.

In recent years, the use of mobile health (mHealth) tools to increase physical activity has risen [7,8]. Furthermore, the widespread use of mobile phones has made mHealth interventions scalable to a broad population [9,10]. Although there are a number of different mHealth tools and services available, smartphone apps may be a particularly popular approach to increasing physical activity.

In 2017, there were >325,000 health apps available from major app stores and approximately 3.7 billion app downloads worldwide [11]. The most popular health apps tend to be for diet, physical activity tracking, weight management, and adherence to medication [7,12,13]. Smartphone apps are generally considered easy to use and can enhance physical activity interventions through technological features (eg, accelerometers) [9]. Moreover, apps have been shown to be effective at increasing physical activity levels [10,14], although there is substantial variability in quality and effectiveness between the many available apps [15,16]. Physical activity apps also tend to be inexpensive or free of charge [10]. For example, in NZ, the Ministry of Health–supported web-based Health Navigator app library contains a number of different mHealth apps and specifically includes links to free and low-cost physical activity apps [17].

The prescription of physical activity apps during a primary care visit is a plausible intervention in the NZ context, as some general practitioners (GPs) already *prescribe* exercise as part of a green prescription program [18], although it is unclear whether there is substantial uptake of the program, and there are no requirements for physical activity levels to be assessed as part of standard care by GPs. Such a program could theoretically include smartphone app prescriptions. Clinicians and GPs already frequently recommend apps and other web-based resources during consultations [19,20], and the Royal New Zealand College of General Practitioners supports the adoption of such technology [21]. In addition, some NZ GPs recommend pedometer use for certain patients [22] and would presumably recommend the mHealth equivalent.

Given this background, this study assesses the health impacts, health system costs, and cost-effectiveness of the prescription of smartphone apps for the promotion of physical activity in primary care settings in NZ, a high-income country.

Methods

Modeling Methods

An established proportional multistate life table (PMSLT) model was used to estimate the health impact and health system expenditure of the prescription of smartphone apps for the promotion of physical activity in primary care [23,24]. Physical activity was measured as the change in the metabolic equivalent of task (MET) minutes per week of moderate and vigorous activity. The PMSLT model simulates the entire NZ population alive in 2011 (N=4.4 million) until death or the age of 110 years. Future all-cause morbidity and mortality and incidence and case fatality rates for 5 diseases related to changes in physical activity were projected. Specifically, the model included breast cancer (women only), colorectal cancer, CHD, type 2 diabetes, and stroke.

Health gain was measured in quality-adjusted life years (QALYs) [25], whereas, for costs, a health system perspective was used, and the outputs were the difference in total health system costs between business-as-usual and the modeled intervention and included the cost of implementing the intervention. We also disaggregated the results by period and presented the impact of the intervention after 10 years and 20 years. Calibration and validation of the epidemiological aspects of the PMSLT are described in a web-based technical report [23]. A Monte Carlo simulation (2000 iterations) was used to estimate the uncertainty intervals for the key results.

QALYs and costs were discounted at 3%, with results for 0% and 6% discount rates presented as scenario analyses. We also ran the results applying an *equity adjustment* that set background all-cause morbidity and mortality rates for Māori to non-Māori values [26]. This technique is often used to avoid the undervaluation of health gains and identify potential health equity impacts for Māori. Scenario analyses included higher percentages of the eligible population being screened for the intervention (ie, 25% and 50%), a reversed ratio of GP to practice nurse (PN) consultation time, and maintenance of the intervention impact for 5 years.

Full details of the model are published elsewhere [23,24,27].

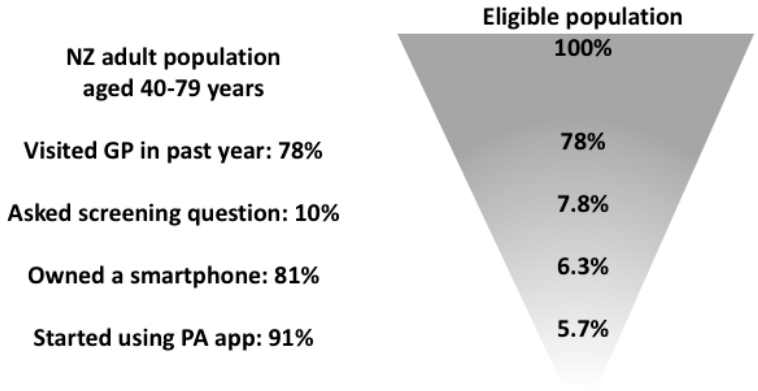
Intervention Specification

Rapid reviews of the literature on physical activity apps and referral schemes (ie, green prescription programs) were conducted to parameterize the intervention for modeling purposes. The modeling parameters used and the justification for their use are presented below (Tables 1 and 2).

As shown in Figure 1, during a GP visit, people with insufficient physical activity were identified using a screening question as per the one used by the NZ Health Survey [28]. For practical reasons and because it might be inappropriate to ask on some occasions, the screening question was assumed to only be asked during a proportion of such visits (10%, with 25% and 50% presented as scenario analyses). For those with insufficient levels of physical activity (defined as <300 MET minutes per week), the GP or PN then provided a printed physical activity prescription form that included specific instructions to exercise and information on how to download and use a physical activity smartphone app. The patient then chose a physical activity app, possibly with input or recommendations from the provider, from the Health Navigator app library. The Health Navigator app library is a repository of apps reviewed by experts, which contains links to a number of independently reviewed exercise apps [17]. It was assumed that individuals would use an app that was already developed and free to download. Furthermore, it was assumed that the app would be available at zero cost for as long as the participants chose to use it. However, adherence to the intervention was expected to decrease relatively quickly;

therefore, long-term availability of the app would not be relevant for most participants. In addition to the prescription, GPs or PNs referred patients to service providers (eg, a regional sports trust or primary health organization in the NZ context) who provided support over the phone, including a comprehensive consultation, 2 brief follow-up calls, and technical support to use the app within the first year after the initial consultation and prescription.

Figure 1. Flowchart of base-case intervention conceptualization for prescribed smartphone apps for physical activity promotion in primary care. GP: general practitioner; NZ: New Zealand; PA: physical activity.



We included direct costs attributable to the intervention (ie, intervention costs) and indirect costs attributable to changes in health system use resulting from the intervention [29]. Intervention costs were applied to everyone who received an initial consultation, even if they did not start using a physical activity app. In addition, it was assumed that all costs would be incurred in the first year following the initial consultation (ie, follow-up calls and technical support would occur in the same year as the intervention). Furthermore, it was assumed that those adults who were sufficiently active or who did not own a smartphone would require less than a minute of the GP's time to screen; therefore, costs associated with such interactions were minimal and not quantified.

Current evidence indicates that physical activity interventions would likely provide minimal chronic disease reduction benefits to adults aged <40 years in NZ [24,27]. Therefore, the age range for the intervention was restricted to 40 to 79 years. Individuals aged ≥80 years were excluded because of low smartphone ownership and use, as well as the relatively high prevalence of comorbidities (eg, arthritis) that could limit participation in physical activity. Although the intervention was only applied to those aged 40 to 79 years, the model followed the entire NZ population alive in 2011 over their remaining life span. Therefore, participants could age into the intervention over time (eg, someone who was aged 38 years in 2011 could still become eligible for the intervention in 2013 when they turned 40).

On the basis of evidence in the literature (Table 1), it was assumed that 91% of eligible patients who were active smartphone users would start the intervention and that physical activity would increase by an average of 410 moderate to vigorous physical activity MET minutes per week. After 1 year, the intervention effect was assumed to be maintained by 36.5% of people, with a decay rate of 55% applied.

Parameters

Tables 1 and 2 contain additional details on the base-case intervention parameters.

Table 1. Modeling input parameters for the prescription of smartphone apps for physical activity promotion in primary care.

Parameter and key source	Supporting evidence and notes	Value (UI[a]; beta distribution unless otherwise indicated)	Resulting percentage (alternate scenarios)
Visited GP[b] in the past year			
Ministry of Health [30]	According to the NZ[c] Health Survey (NZHS), 78% of NZ adults (aged ≥15 years) visited their GP in the past year [30]. The uncertainty intervals are an assumed percentage (±10%).	78% (68%-88%)	78%
Asked screening question			
Croteau et al [31]	A national survey on physical activity and nutrition in NZ by Croteau et al [31] found that only 3% of the survey population (n=235) reported receiving a green prescription from a GP or PN[d]. In addition, the study reported that those aged ≥45 years were significantly more likely to have received a green prescription. On the basis of the findings from Croteau et al [31] and survey data from the NZHS, which indicated that 49% of people who visit their GP have insufficient levels of physical activity [30], it was back-calculated that only approximately 10% of the eligible population would be asked the physical activity screening question during any of their GP visits within a year (ie, the estimate for NZ adults who visited their GP in the previous year [78%] multiplied by the estimate for those asked the screening question [10%], the estimate by Croteau et al [31] for insufficient physical activity [49%], and uptake of the app [81%] equals 3%, which is the percentage of the eligible population that would be expected to receive a green prescription from a GP or PN). However, several alternate scenarios with higher percentages of the eligible population being asked the screening question are also considered in scenario analysis. Although no other evidence was found to support parameter selection, incomplete screening was still included for several reasons. For example, the patient may not be able to exercise because of an existing health condition. In addition, it may be inappropriate to ask the screening question during a consultation about an urgent and critical other matter, and the GP may not have the time for noncritical care provision.	10% per year (alternate scenario 1: 25%; alternate scenario 2: 50%)	7.8% of the eligible population (alternate scenario 1: 19.5%; alternate scenario 2: 39%)
Smartphone ownership			
DataReportal [32]	Using recent metrics on NZ smartphone ownership (based on Google Consumer Barometer data), it has been reported that 81% of NZ adults own a smartphone [32]. The uncertainty intervals are an assumed percentage (±5% of the point estimate).	81% (77%-85.1%)	6.3% of the eligible population (alternate scenario 1: 15.8%; alternate scenario 2: 31.6%)
Uptake of the smartphone app			
Glynn et al [33]	On the basis of an RCT[e] of physical activity apps prescribed in primary care in Ireland, 91% of eligible patients who were active smartphone users were assumed to start the intervention [33]. The uncertainty intervals are an assumed percentage (±10% of the point estimate).	91% (81.9%-100%)	5.7% of the eligible population (alternate scenario 1: 14.4%; alternate scenario 2: 28.8%)
Increase in physical activity			
Glynn et al [33]	On average, mHealth[f] physical activity interventions result in an increase in physical activity, at least in the short term [10,14,34]. As a result of the intervention, it was assumed that physical activity would increase on average by 410 MVPA[g] MET[h] minutes per week. This total was taken from an RCT studying a GP-prescribed physical activity app and follow-up support [33]. The study reported a 2017-step increase per day after accounting for differences between the intervention and control groups. Steps per day were converted to MVPA MET minutes per week using the method outlined in a web-based report [35]. The *Steps to MVPA Conversion* section of this report also has further details about the formula used. The uncertainty intervals are an assumed percentage (±10% of the point estimate).	Increase by 410 (369-451) minutes of MVPA MET minutes per week; normal distribution	—[i]
Adherence at 1 year			

Parameter and key source	Supporting evidence and notes	Value (UI[a]; beta distribution unless otherwise indicated)	Resulting percentage (alternate scenarios)
Allman-Farinelli et al [36] and Damschroder et al [37]	After 1 year, the intervention effect was assumed to be maintained by 36.5% of people. This was based on the average of estimates for 2 app+ (ie, a smartphone app in addition to follow-up texts, calls, or emails) intervention studies (see below), which typically fall between the estimates for traditional green prescription programs and app-only physical activity interventions in retention and adherence. The uncertainty intervals are an assumed percentage (±20% of the point estimate). The first RCT for the prevention of weight gain in young adults in Australia found that an app+ intervention that targeted both dietary behaviors and physical activity generated a 40% response rate to follow-up SMS text messages at 9 months [36]. The second RCT for physical activity in US veterans reported a 33% retention rate at 12 months for an app+ group that received follow-up phone calls [37].	36.5% (29.2%-43.8%)	—
Decay rate of intervention effect after 1 year			
Gc et al [38]	A recent modeling study of brief physical activity interventions also used a similar methodological approach and assumed that the interventions had an effect for the first year and then applied a 55% decay rate every year afterward [38]. This was in line with several previously reported physical activity modeling studies (ie, Over et al [39], Cobiac et al [40], and Jacobs-van der Bruggen et al [41]) that assumed similar base-case decay rates, varying between 50% and 55%. The uncertainty intervals are an assumed percentage (±20%).	55% (35%-75%)	—

[a]UI: uncertainty interval.

[b]GP: general practitioner.

[c]NZ: New Zealand.

[d]PN: practice nurse.

[e]RCT: randomized controlled trial.

[f]mHealth: mobile health.

[g]MVPA: moderate to vigorous physical activity.

[h]MET: metabolic equivalent of task.

[i]Not available (does not change % of eligible population).

Table 2. Cost input parameters for the prescription of smartphone apps for physical activity promotion in primary care.

Parameter	Key source	Supporting evidence and notes	Value (95% UI[a])
Ratio of GP[b] to PN[c] consultations	Research New Zealand [42]	Approximately 73% of consultations were assumed to be GP-run and the rest were run by PNs. These proportions are based on the referral sources reported by the NZ[d] Green Prescription Patient Survey [42].	73% GP, 27% PN; however, in a scenario analysis, this ratio was reversed.
GP consultation parameters			
GP consultation time	Elley et al [43]	On the basis of an RCT[e] studying the NZ Green Prescription Program [43], 7 minutes of GP time were spent on the physical activity advice and prescription part of each consultation. Although it is likely that the overall consultation will typically be longer, only the physical activity–specific part has been quantified. Other studies have reported a longer duration [38]; however, NZ-specific data were used for this parameter. It was assumed that there would be additional GP time available for the intervention and, therefore, all other patient concerns would still be covered in the appointment, and no adverse effects would arise from the GP consultation. The uncertainty intervals are an assumed percentage (±10% of the point estimate).	7 minutes (6.3-7.7)
Cost of GP consultation in 2011	Association of Salaried Medical Specialists [44]	The cost of 7 minutes of a GP consultation was assumed to be NZ $15.38 (US $10.35), or NZ $2.20 (US $1.48) per minute. The midpoint of a GP annual salary scale in 2018 was taken from the Wellington Union Health Services Collective Agreement [44]. An hourly rate of NZ $94.15 was then calculated using this estimate. With 50% overheads, this equates to NZ $141.20, or NZ $131.85 in 2011 adjusted for inflation [45]. For 7 minutes of a consultation at an hourly rate of NZ $131.85, the physical activity part of the GP consultation would cost NZ $15.38 per consultation. By comparison, the NZ government agency PHARMAC estimated that the cost of a GP practice visit was NZ $80 per consultation in 2018 [46]. This equates to NZ $73.73 per consultation in 2011 [45], and NZ $33.73 once a patient copayment of NZ $40 is removed. If 7 out of 15 minutes were allocated to physical activity advice and prescription, the cost would be approximately NZ $15.74, which is close to the estimate above. The final costs have been presented as 2011 NZ $. The baseline year of the model was 2011, and cost parameters were consumer price index–adjusted to the 2011 NZ $ to reflect this. With the exception of costs, other parameters in this table are more current, so they produce more relevant outputs.	NZ $2.20/minute (US $1.48/minute)
PN consultation parameters			
PN consultation time	Elley et al [43]	A PN was assumed to spend approximately 13 minutes on the physical activity app consultation based on the results of an RCT on the NZ Green Prescription Program [43]. It was assumed that there would be additional PN time available for the intervention and, therefore, all other patient concerns would still be covered in the appointment, and no adverse effects would arise from the consultation. The uncertainty intervals are an assumed percentage (±10% of the point estimate).	13 minutes (11.7-14.3)
Cost of PN consultation in 2011	Elley et al [47]	The cost of a 13-minute consultation was assumed to be NZ $8.27 (2011 US $5.57), consumer price index–adjusted to the 2011 NZ $, or NZ $0.64 (US $0.43) per minute. A PN hourly wage was NZ $19.12/hour in 2000-2001 [47], equivalent to NZ $25.42 in 2011 [45]. This equates to NZ $38.16 per hour with overheads (as per the GP calculations) and NZ $8.27 for a 13-minute consultation. By comparison, the midpoint of the Practice Nurse Collective Employment Agreement pay scale was NZ $24.36 per hour [48]. This is similar to our hourly rate before adjusting for overheads.	NZ $0.64/minute (US $0.43/minute)
Additional costs[f]			

Parameter	Key source	Supporting evidence and notes	Value (95% UI[a])
Cost of additional resources in 2011	Elley et al [47]	The cost of follow-up phone calls and additional resources was assumed to be NZ $90.10 (US $60.63) per individual. As per the structure of the NZ Green Prescription Program, the intervention was assumed to also include phone calls and additional resource use. After consultation, the intervention would include 3 follow-up phone calls, the first a comprehensive consultation and then 2 brief follow-up calls. The phone calls would include general advice on physical activity and technical support to use the app. Additional resources would include educational material dissemination, such as an email with a link to a website with responses to frequently asked questions. Similar services were estimated to cost NZ $69 per person in 2001-2002 based on the results of a PhD thesis on the NZ Green Prescription Program [47]. This NZ $69 in 2001 equates to NZ $90.10 (US $60.63) in 2011 [45]. The uncertainty intervals are an assumed percentage (±10% of the point estimate).	NZ $90.10 (US $60.63; 81.09-99.11)

[a]UI: uncertainty interval.

[b]GP: general practitioner.

[c]PN: practice nurse.

[d]NZ: New Zealand.

[e]RCT: randomized controlled trial.

[f]It was assumed that individuals would use an app that was already developed and was free to download from the Health Navigator website (ie, zero cost for the app); it was also assumed that there was zero cost for promoting the app to primary care workers.

Results

The prescription of smartphone apps for physical activity promotion in primary care resulted in an increase of 430 QALYs (95% uncertainty interval 320-550) over the lifetime of the 2011 NZ population. This was equivalent to 0.13 QALYs gained per 1000 population (95% uncertainty interval 0.01-0.16; Table 3) or 0.23 QALYs gained per 1000 adults aged 40 to 79 years. Of the total, 150 QALYs and health system cost savings of NZ $174,000 (US $117,000) were accumulated in the first 10 years following intervention implementation, and 160 QALYs and cost savings of NZ $1,961,000 (US $1,320,000) were accumulated in the 20 years following implementation.

The modeled improvements in health came with net cost savings of NZ $2.2 million (US $1.5 million). The intervention was cost-saving for all age-sex-ethnicity subgroups, except for non-Māori women aged 40 to 59 years.

On a per capita basis, QALY gains were generally larger in women than in men, larger in Māori than in non-Māori, and largest in the 60 to 79 years age group. Health gains for Māori increased with the application of the equity adjustment (ie, non-Māori morbidity and mortality rates used for Māori; Table 4).

In the first 10 years after the intervention was implemented (2011-2020), the total health gain was 148 QALYs, with net cost savings of NZ $174,000 (US $117,000) for the health system. After 20 years (2011-2030), the total health gain was 158 QALYs, with net health system cost savings of NZ $1.96 million (US $1.32 million).

The impact of selected changes to model specifications on the results was explored (Table 5). Changing the discount rate had the expected impact on the overall results, with a 6% discount resulting in smaller health gains and cost savings and a 0% discount (ie, undiscounted) resulting in larger health gains and cost savings. Increasing the percentage of primary care patients who were asked the screening question for the intervention also had the expected impact on the overall results, with higher screening rates resulting in higher health gains. However, the cost savings were larger for the scenario in which 25% of patients were asked the screening question (ie, cost savings of NZ $3.3 million [US $2.2 million]) than for the scenario in which 50% were asked (ie, cost savings of NZ $2.6 million [US $1.8 million]) because of the cost of intervention implementation. Dominant provision by PNs (reversing the ratio of GP to PN consultation provision) resulted in a small increase in cost savings (NZ $148,000 [US $100,000]) over the base case scenario. Assuming that the intervention impact would be maintained for 5 years following the intervention (rather than for 1 year), the health gains were estimated to be >4 times higher than in the base-case analysis and would result in much higher cost savings (NZ $22.5 million [US $15.1 million]).

Finally, we examined the contribution of individual intervention parameters to the uncertainty in the modeled results. Uncertainty in health gains and health system cost impacts was primarily driven by the decay rate and the uptake of the smartphone app (Figures 2 and 3).

Table 3. Health gains and health system costs of the prescription of smartphone apps for physical activity promotion in primary care by age, sex, and ethnicity (lifetime gains and 3% discount rate). 2011 NZ $1=2011 US $0.67.

Sex, ethnicity, and age group	Health gain, QALYs[a] (95% UI[b])	QALYs/1000 population (95% UI)	Health system costs, 2011 NZ $ million/2011 US $ million (95% UI)
All sexes and all ethnicities			
All age groups	430 (320 to 550)	0.13 (0.10 to 0.16)	−2.16[c]/−1.45 (−4.49 to −0.11)
40-79 years	430	0.23	−2.16 /−1.45
Male			
Non- Māori			
40-59 years	69 (50 to 89)	0.13 (0.10 to 0.17)	−0.34/−0.23 (−0.83 to 0.09)
60-79 years	103 (75 to 140)	0.34 (0.25 to 0.45)	−0.79/−0.53 (−1.38 to −0.28)
Māori			
40-59 years	19 (14 to 25)	0.30 (0.22 to 0.39)	−0.17/−0.11 (−0.27 to −0.08)
60-79 years	9 (7 to 12)	0.42 (0.31 to 0.55)	−0.09/−0.06 (−0.14 to −0.05)
Female			
Non-Māori			
40-59 years	58 (43 to 76)	0.11 (0.08 to 0.14)	0.17/0.12 (−0.28 to 0.60)
60-79 years	130 (96 to 170)	0.41 (0.30 to 0.54)	−0.69/−0.46 (−1.30 to −0.13)
Māori			
40-59 years	24 (17 to 31)	0.33 (0.24 to 0.42)	−0.18/−0.12 (−0.30 to −0.06)
60-79 years	14 (10 to 18)	0.56 (0.41 to 0.73)	−0.08/−0.05 (−0.14 to −0.03)

[a]QALY: quality-adjusted life year.

[b]UI: uncertainty interval.

[c]Negative cost (ie, the intervention results in cost savings to the health system).

Table 4. Results for Māori (Indigenous population) with equity adjustment applied (40-79 age group, lifetime gains, and 3% discount rate).

Sex and age group	Health gain, QALYs[a] (95% UI[b])	QALYs/1000 population (95% UI)	Health system costs, NZ $ million/US $ million (95% UI)
Male			
40-59 years	24 (18 to 31)	0.37 (0.28 to 0.49)	−0.17/−0.11 (−0.28 to −0.09)
60-79 years	13 (10 to 17)	0.59 (0.44 to 0.77)	−0.09/−0.06 (−0.15 to −0.05)
Female			
40-59 years	29 (22 to 38)	0.40 (0.30 to 0.51)	−0.18/−0.12 (−0.32 to −0.08)
60-79 years	20 (14 to 25)	0.78 (0.58 to 1.01)	−0.08/−0.06 (−0.15 to −0.03)

[a]QALY: quality-adjusted life year.

[b]UI: uncertainty interval.

Table 5. Sensitivity and scenario analyses for the prescription of smartphone apps for physical activity promotion in primary care (expected value analysis, lifetime perspective, and 3% discount rate unless otherwise noted).

Sensitivity and scenario analyses	Health gains, QALYs[a]	Net health system costs, NZ $ million (US $ million)	Cost per QALY gained, NZ $
Base-case analysis	430	−2.162 (−1.455)	Cost saving[b]
Undiscounted	720	−3.820 (−2.571)	Cost saving
6% discount rate	290	−0.900 (−0.605)	Cost saving
25% asked screening question	950	−3.339 (−2.247)	Cost saving
50% asked screening question	1640	−2.644 (−1.779)	Cost saving
Dominant provision by PNs[c] (reversed ratio of GP[d] to PN consultations)	430	−2.310 (−1.555)	Cost saving
5-year maintenance of additional physical activity levels followed by a return to preintervention levels (otherwise base case)	1750	−22.490 (−15.135)	Cost saving

[a]QALY: quality-adjusted life year.

[b]Negative cost per QALY gained (ie, the intervention results in cost savings to the health system).

[c]PN: practice nurse.

[d]GP: general practitioner.

Figure 2. Tornado plot showing the contribution of parameter uncertainty to overall uncertainty in the quality-adjusted life years gained for the studied adult population. GP: general practitioner; PN: practice nurse; QALY: quality-adjusted life year.

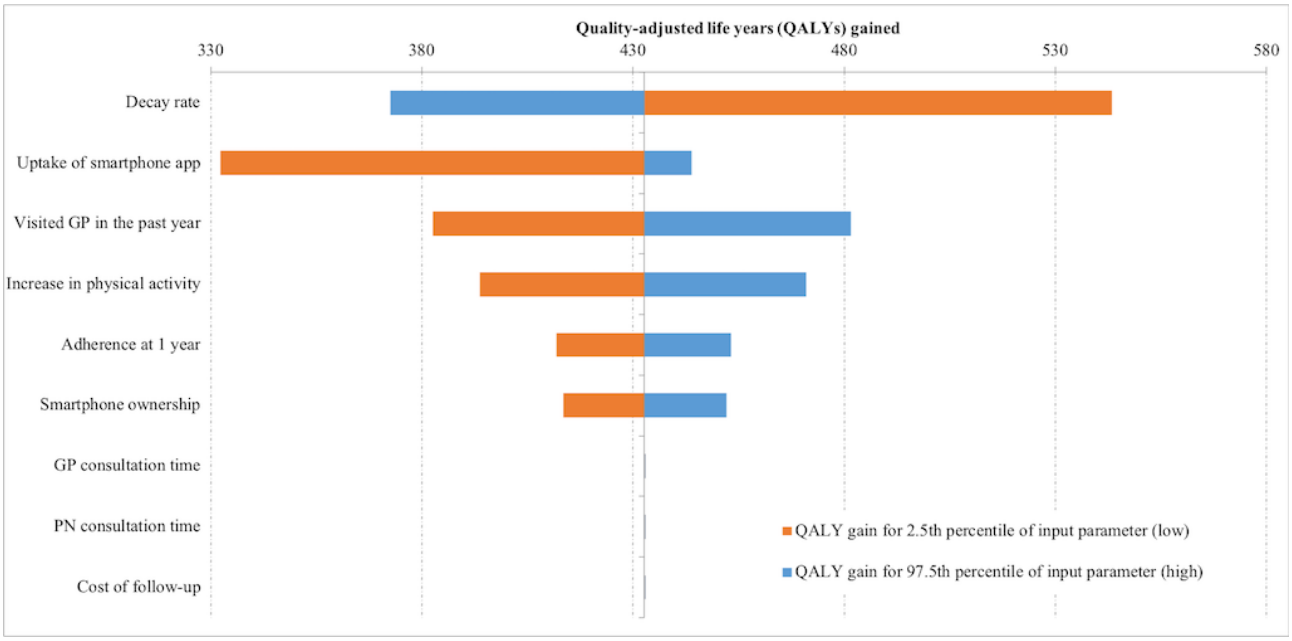
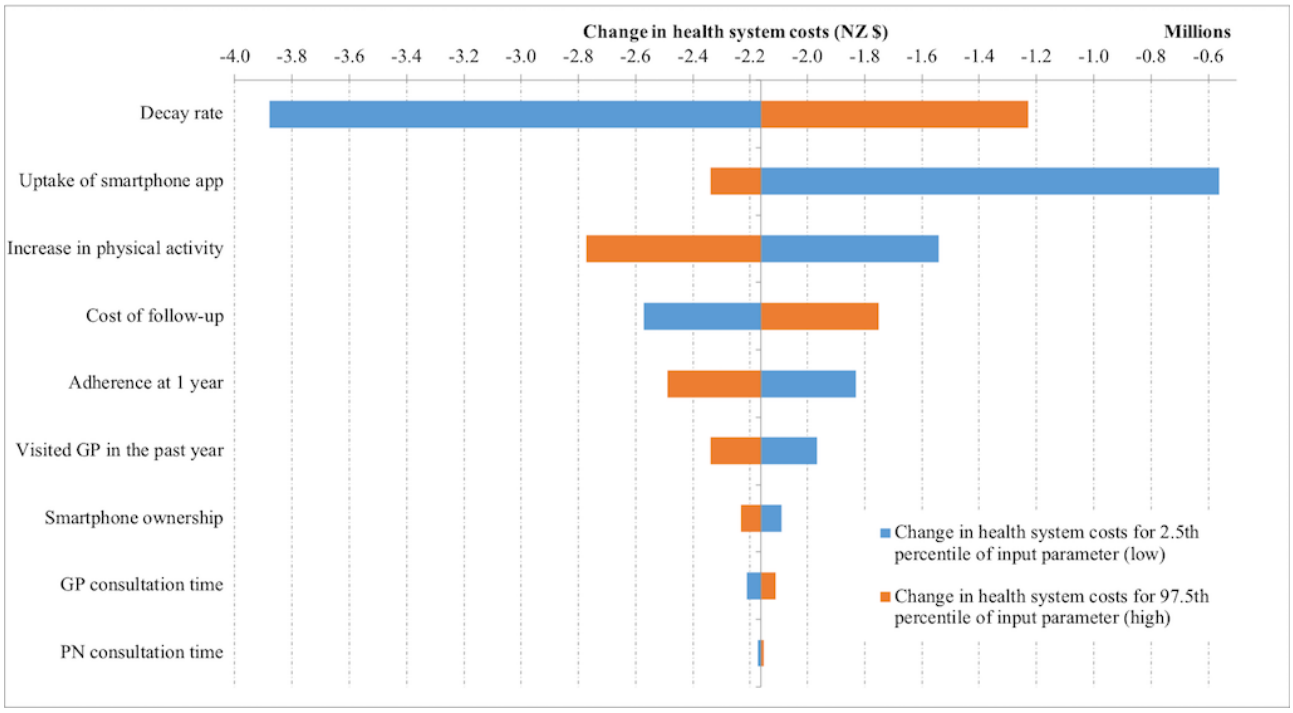


Figure 3. Tornado plot showing the contribution of parameter uncertainty to overall uncertainty in the change in health system costs. GP: general practitioner; PN: practice nurse. 2011 NZ $1=2011 US $0.67.



Discussion

Principal Findings and Interpretation

The potential impact of the prescription of smartphone apps for physical activity promotion in primary care settings in NZ was modeled using published estimates of uptake, effectiveness, and adherence [33,36,37]. The total health impact was modest, with 430 QALYs gained over the remaining life span of the population, albeit up to 1750 QALYs if the intervention effect was maintained for 5 years. The intervention was also likely to be cost saving for the health system and provide larger per capita health gains for Māori than for non-Māori. Given the higher per capita health gains for Māori than for non-Māori, it is plausible that this intervention could play a role in reducing health inequalities if implemented equitably. Although there is strong evidence that physical activity levels can be improved and maintained in the short term through individual-level physical activity interventions, a cumulative meta-analysis of the effects of physical activity interventions found that additional research is needed to identify interventions that are the most cost-effective [49]. This study provides additional modeling evidence of the cost-effectiveness of physical activity promotion in a primary care setting.

Comparison With Prior Work

When compared with selected health interventions in NZ, according to methodologically compatible modeling studies, we found that the prescription of smartphone apps for physical activity promotion in primary care was likely to provide larger health gains and cost savings for the health system than a mass media campaign for physical activity apps [27], a mass media campaign for weight loss apps [50], or weight loss counseling by nurses in primary care [51] in NZ (Table 6). However, the intervention was less effective than a mass media campaign to promote a smoking cessation app in NZ [52]. It was also less effective on a per capita basis than a traditional *green prescription* program to promote physical activity in Australia [40], although this may be due in part to underlying differences in physical activity patterns and epidemiology across different populations. We also note that the health gains of these individual-level interventions are orders of magnitude lower than upstream interventions (eg, tobacco control endgame interventions [53] and switching driving trips to walking and cycling [24]). Implementing the prescription of smartphone apps for physical activity promotion in primary care alongside other such interventions may help maximize health gains.

Table 6. Comparison of the impact of various health interventions in New Zealand according to methodologically compatible epidemiological and health economic modeling (lifetime perspective and 3% discount rate).

Intervention	Health gains, QALYs[a]	Net health system costs, NZ $ million[b]	Cost per QALY gained (incremental cost-effectiveness ratio), NZ $[b]
Prescription of smartphone apps for physical activity promotion in primary care (this study)	430	−2.2	Cost saving
Mass media campaign to promote physical activity apps [27]	28	2.2	81,000
Mass media campaign to promote weight loss apps [50]	29	2.9	79,700
Weight loss dietary counseling by nurses in primary care [51]	250	38.8	138,000
5-year mass media campaign to promote smoking cessation app [52]	6760	−115.0	Cost saving
Enhanced green prescription program among women aged 40-74 years [54]	—[c]	—[c]	687[d]

[a]QALY: quality-adjusted life year.

[b]2011 NZ $1=2011 US $0.67.

[c]This study did not use the same modeling approach but calculated cost-effectiveness ratios.

[d]Program cost per person made *active* and sustained at 12 months.

Study Strengths and Limitations

A strength of this study was the use of an established PMSLT model for physical activity [24,27], which is based on high-quality disease-specific epidemiological and costing data for a whole country. The modeling framework has been widely used in Australia and NZ to assess different health interventions, including several different mHealth interventions [27,50,52], allowing for comparisons across studies. However, a limitation of the PMSLT approach is the assumption of disease independence, although the model does account for the relationship between type 2 diabetes and CHD and stroke, given that type 2 diabetes is a risk factor for these conditions. The model also does not account for potentially beneficial health impacts via the reduction of obesity, anxiety, or depression (or, conversely, health loss via an increase in these conditions because of musculoskeletal injuries associated with physical activity). Another limitation of this study was the use of a health system perspective for costs and benefits. Such a perspective does not allow for the estimation of societal-level costs and benefits outside of the health system, such as impacts on household spending or reduced greenhouse gas emissions from transport mode shifts.

There are also limitations associated with the parameter estimates used. For example, app uptake likely varies and could be increased with improved app design. In addition, adherence estimates in the literature range widely. We examined estimates from green prescription programs, stand-alone app interventions, and app+ interventions (ie, a smartphone app in addition to follow-up texts, calls, or emails). We chose an adherence estimate between those for green prescription programs and app-only interventions, which was in line with app+ interventions in which there would be some level of follow-up with the intervention participants. We assumed that all of the intervention parameters presented in Tables 1 and 2, including intervention uptake and adherence, would be the same across population groups (ie, men and women and Māori and non-Māori). However, it is possible that for certain parameters, there may be significant variations across population groups that may impact the effectiveness of the intervention. For example, it is unclear whether currently available apps adequately cater to the needs of diverse population groups.

Potential Implications for Research

Additional research is needed to optimize interventions that can lead to sustained increases in physical activity levels over the long term [48] and identify the forms of follow-up (eg, phone calls vs emails) that maximize adherence to physical activity interventions. App improvements may also encourage adherence. Evidence suggests that apps are most effective when they incorporate self-monitoring of physical activity, reminders for app use, and social interaction with peers [7,55]. Gamification (eg, providing point-based rewards for frequency or consistency of app use) may also improve outcomes for app-based interventions [56]. Other factors that may influence app use include simple app interfaces, easy navigation, and automatic or simplified data entry [57]. Such features help increase user engagement, which promotes positive behaviors [7]. Advancements in app design merit additional research to assess their impact on adherence.

Potential Implications for Policy

There are indications that such a program would be best administered by PNs rather than GPs, given the larger cost savings associated with the scenario in which PNs are the dominant deliverers of the consultation. In addition, the literature suggests that GPs are often particularly time-limited [58]. Our scenario analyses suggest that larger health gains would be achievable if a higher percentage of patients were asked the screening question (ie, 25% or 50% asked the screening question). If GPs are unable to ask the screening question as frequently as PNs because of time constraints, then it may be both more effective and a better use of staff resources to routinely have PNs deliver such a program. This may also have implications for the administration of the ongoing green prescription program in NZ. However, the ratio of GP to PN consultations was only included in the model as a cost parameter, and we did not assess potential differences in effectiveness between GP and PN administration of the program.

We also note that, in the context of high levels of unmet need for health care (eg, 13.3% of NZ adults did not visit a GP because of cost barriers in 2019-2020 [59]) and differential access to smartphones and the internet, additional strategies to ensure that any app-based intervention does not exacerbate existing health inequities are warranted.

With the widespread use of smartphones, mHealth interventions such as this have a large potential for scalability to a broad population [9,10]. As part of a range of interventions to address insufficient physical activity, governments should consider investing in the promotion of physical activity smartphone apps, along with additional research to improve app effectiveness and uptake. The intervention should ideally be tailored to the country context, which may include examining existing structural inequities that may influence intervention success and co-designing strategies with relevant population groups.

The promotion of smartphone apps may also complement other strategies to promote physical activity (eg, investments in walking and cycling infrastructure [60]) to support long-term behavioral changes. Although our results suggest that the promotion of physical activity smartphone apps in primary care is likely to be effective and cost-saving in NZ, these results are also likely generalizable to other high-income countries with similar chronic disease epidemiology, physical activity levels, and other population characteristics.

Conclusions

In this modeling study, the prescription of smartphone apps for physical activity promotion in primary care in NZ yielded modest health gains and was cost saving for the health care system. The scope for this type of mHealth intervention is expanding with the increase in smartphone ownership and the availability of easy-to-use and effective apps. This intervention should be considered by policy makers in NZ and also be considered by other high-income nations with similar characteristics.

Acknowledgments

The authors gratefully acknowledge Tony Blakely and Linda Cobiac, who helped with other aspects of building the physical activity model used in this study. This study was funded by the Health Research Council of New Zealand (program grants 13/724 and 16/443). The funder had no role in the study design, analysis or interpretation of the data, writing of the report, or the decision to submit the paper for publication.

Conflicts of Interest

None declared.

References

1. Wahid A, Manek N, Nichols M, Kelly P, Foster C, Webster P, et al. Quantifying the association between physical activity and cardiovascular disease and diabetes: a systematic review and meta‐analysis. J Am Heart Assoc 2016 Sep;5(9):e002495. [doi: 10.1161/jaha.115.002495]
2. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. Br Med J 2016 Aug 09;354:i3857 [FREE Full text] [doi: 10.1136/bmj.i3857] [Medline: 27510511]
3. Physical activity. World Health Organization. 2020 Nov 26. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2021-06-30]
4. Global Health Observatory Data Repository: prevalence of insufficient physical activity among adults. World Health Organization. 2018. URL: https://apps.who.int/gho/data/view.main.2463?lang=en [accessed 2021-06-30]
5. Health loss in New Zealand: a report from the New Zealand burden of diseases, injuries and risk factors study, 2006-2016. Ministry of Health. 2013. URL: https://www.moh.govt.nz/notebook/nbbooks.nsf/0/F85C39E4495B9684CC257BD3006F6299/$file/health-loss-in-new-zealand-final.pdf [accessed 2021-06-30]
6. Pylypchuk R, Wells S, Kerr A, Poppe K, Riddell T, Harwood M, et al. Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. Lancet 2018 May;391(10133):1897-1907. [doi: 10.1016/s0140-6736(18)30664-0]
7. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]
8. Paglialonga A, Lugo A, Santoro E. An overview on the emerging area of identification, characterization, and assessment of health apps. J Biomed Inform 2018 Jul;83:97-102 [FREE Full text] [doi: 10.1016/j.jbi.2018.05.017] [Medline: 29852318]
9. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
10. Feter N, Dos Santos T, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019 Jun 13;64(5):679-690. [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]

11. mHealth app economics 2017/2018: current status and future trends in mobile health. Reasearch2Guidance. 2017. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2021-06-30]
12. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]
13. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
14. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
15. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
16. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
17. About the health navigator app library. Health Navigator New Zealand. 2019. URL: https://www.healthnavigator.org.nz/apps/p/people-process/ [accessed 2021-06-30]
18. How the green prescription works. Ministry of Health. 2017. URL: https://www.health.govt.nz/our-work/preventative-health-wellness/physical-activity/green-prescriptions/how-green-prescription-works [accessed 2021-06-30]
19. Boyle L, Grainger R, Hall RM, Krebs JD. Use of and beliefs about mobile phone apps for diabetes self-management: surveys of people in a hospital diabetes clinic and diabetes health professionals in New Zealand. JMIR Mhealth Uhealth 2017 Jun 30;5(6):e85 [FREE Full text] [doi: 10.2196/mhealth.7263] [Medline: 28666975]
20. 2017 HINZ conference proceedings. Health Informatics New Zealand. 2017. URL: https://cdn.ymaws.com/hinz.site-ym.com/resource/resmgr/conference_2017/HiNZ17_proceedings.pdf [accessed 2021-12-02]
21. Telehealth and technology-based health services in primary care. The Royal New Zealand College of General Practitioners. 2017. URL: https://www.nzdoctor.co.nz/sites/default/files/2018-07/2017.09-Telehealth-and-technology-based-health-services-in-primary-care.pdf [accessed 2021-06-30]
22. Patel A, Kolt GS, Schofield GM, Keogh JW. General practitioners' views on the role of pedometers in health promotion. J Prim Health Care 2014 Jun 01;6(2):152-156. [Medline: 24892134]
23. Mizdrak A, Blakely T, Cleghorn LL, Cobiac LJ. Technical report for BODE³ active transport and physical activity model. Department of Public Health, University of Otago, Wellington. 2018. URL: https://www.otago.ac.nz/wellington/otago698875.pdf [accessed 2021-06-30]
24. Mizdrak A, Blakely T, Cleghorn CL, Cobiac LJ. Potential of active transport to improve health, reduce healthcare costs, and reduce greenhouse gas emissions: A modelling study. PLoS One 2019 Jul 17;14(7):e0219316 [FREE Full text] [doi: 10.1371/journal.pone.0219316] [Medline: 31314767]
25. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population health. Annu Rev Public Health 2002;23:115-134. [doi: 10.1146/annurev.publhealth.23.100901.140513] [Medline: 11910057]
26. McLeod M, Blakely T, Kvizhinadze G, Harris R. Why equal treatment is not always equitable: the impact of existing ethnic health inequalities in cost-effectiveness modeling. Popul Health Metr 2014;12:15 [FREE Full text] [doi: 10.1186/1478-7954-12-15] [Medline: 24910540]
27. Mizdrak A, Telfer K, Direito A, Cobiac LJ, Blakely T, Cleghorn CL, et al. Health gain, cost impacts, and cost-effectiveness of a mass media campaign to promote smartphone apps for physical activity: modeling study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e18014 [FREE Full text] [doi: 10.2196/18014] [Medline: 32525493]
28. Questionnaires and content guide 2017/18: New Zealand health survey. Ministry of Health. 2018. URL: https://www.health.govt.nz/publication/methodology-report-2017-18-new-zealand-health-survey [accessed 2021-06-30]
29. Kvizhinadze G, Nghiem N, Atkinson J, Blakely T. Cost off-sets used in BODE3 multistate lifetable models. Department of Public Health, University of Otago, Wellington. 2016. URL: https://www.otago.ac.nz/wellington/otago619391.pdf [accessed 2021-06-30]
30. Annual data explorer 2017/18: New Zealand health survey [data file]. Ministry of Health. 2019. URL: https://minhealthnz.shinyapps.io/nz-health-survey-2017-18-annual-data-explorer [accessed 2021-06-30]
31. Croteau K, Schofield G, McLean G. Physical activity advice in the primary care setting: results of a population study in New Zealand. Aust N Z J Public Health 2006 Jun;30(3):262-267. [doi: 10.1111/j.1467-842x.2006.tb00868.x] [Medline: 16800204]
32. Digital New Zealand (January) v01. DataReportal. 2019. URL: https://www.slideshare.net/DataReportal/digital-2019-new-zealand-january-2019-v01 [accessed 2021-06-30]

33. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):384-391 [FREE Full text] [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
34. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
35. Telfer K, Wilson N, Direito A, Mizdrak A. Technical report for BODE3 intervention parameter selection: mobile health for physical activity. Department of Public Health, University of Otago, Wellington. 2020. URL: https://www.otago.ac.nz/wellington/departments/publichealth/research/otago740534.pdf [accessed 2021-06-30]
36. Allman-Farinelli M, Partridge SR, McGeechan K, Balestracci K, Hebden L, Wong A, et al. A mobile health lifestyle program for prevention of weight gain in young adults (TXT2BFiT): nine-month outcomes of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jun 22;4(2):e78 [FREE Full text] [doi: 10.2196/mhealth.5768] [Medline: 27335237]
37. Damschroder LJ, Buis LR, McCant FA, Kim HM, Evans R, Oddone EZ, et al. Effect of adding telephone-based brief coaching to an mhealth app (Stay Strong) for promoting physical activity among veterans: randomized controlled trial. J Med Internet Res 2020 Aug 04;22(8):e19216 [FREE Full text] [doi: 10.2196/19216] [Medline: 32687474]
38. Gc VS, Suhrcke M, Hardeman W, Sutton S, Wilson EC, Very Brief Interventions Programme Team. Cost-effectiveness and value of information analysis of brief interventions to promote physical activity in primary care. Value Health 2018 Jan;21(1):18-26 [FREE Full text] [doi: 10.1016/j.jval.2017.07.005] [Medline: 29304936]
39. Over EA, Wendel-Vos GW, van den Berg M, Reenen H, Tariq L, Hoogenveen RT, et al. Cost-effectiveness of counseling and pedometer use to increase physical activity in the Netherlands: a modeling study. Cost Eff Resour Alloc 2012 Sep 24;10(1):13 [FREE Full text] [doi: 10.1186/1478-7547-10-13] [Medline: 23006466]
40. Cobiac LJ, Vos T, Barendregt JJ. Cost-effectiveness of interventions to promote physical activity: a modelling study. PLoS Med 2009 Jul 14;6(7):e1000110 [FREE Full text] [doi: 10.1371/journal.pmed.1000110] [Medline: 19597537]
41. van der Bruggen MA, Bos G, Bemelmans WJ, Hoogenveen RT, Vijgen SM, Baan CA. Lifestyle interventions are cost-effective in people with different levels of diabetes risk: results from a modeling study. Diabetes Care 2007 Jan;30(1):128-134. [doi: 10.2337/dc06-0690] [Medline: 17192345]
42. Green prescription patient survey 2018 report: Ministry of Health. Research New Zealand. 2019. URL: https://www.health.govt.nz/system/files/documents/publications/green-prescription-patient-survey-2018-report.pdf [accessed 2021-06-30]
43. Elley C, Kerse N, Arroll BL, Robinson E. Effectiveness of counselling patients on physical activity in general practice: cluster randomised controlled trial. Br Med J 2003 Apr 12;326(7393):793 [FREE Full text] [doi: 10.1136/bmj.326.7393.793] [Medline: 12689976]
44. Wellington Union health services collective agreement. Association of Salaried Medical Specialists. 2019. URL: https://www.asms.org.nz/wp-content/uploads/2019/05/FINAL-SIGNED-Wellington-Union-Health-Services-1-January-2019-31-December-2020.pdf [accessed 2021-06-30]
45. Inflation calculator. Reserve Bank of New Zealand. URL: https://www.rbnz.govt.nz/monetary-policy/inflation-calculator [accessed 2021-06-30]
46. Cost resource manual version 3 - October 2018. PHARMAC. 2019. URL: https://www.pharmac.govt.nz/medicines/how-medicines-are-funded/economic-analysis/cost-resource-manual/#s2 [accessed 2021-06-30]
47. Elley CR. The effectiveness and cost-effectiveness of the green prescription physical activity intervention: A cluster randomised controlled trial in primary health care. In: Thesis and Dissertations - University of Auckland. Auckland: University of Auckland; 2003.
48. Primary health care multi-employer collective agreement. New Zealand Nurses Organisation. 2012. URL: https://assets-global.website-files.com/5db268b46d028bbc0fc0b537/5de327ea3719a1cf0ee63227_PHC-MECA-2012-2014-signed.pdf [accessed 2021-06-30]
49. Love R, Adams J, van Sluijs EM, Foster C, Humphreys D. A cumulative meta-analysis of the effects of individual physical activity interventions targeting healthy adults. Obes Rev 2018 Aug 27;19(8):1164-1172 [FREE Full text] [doi: 10.1111/obr.12690] [Medline: 29701299]
50. Cleghorn C, Wilson N, Nair N, Kvizhinadze G, Nghiem N, McLeod M, et al. Health benefits and cost-effectiveness from promoting smartphone apps for weight loss: multistate life table modeling. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e11118 [FREE Full text] [doi: 10.2196/11118] [Medline: 30664471]
51. Cleghorn C, Wilson N, Nair N, Kvizhinadze G, Nghiem N, McLeod M, et al. Health benefits and costs of weight-loss dietary counselling by nurses in primary care: a cost-effectiveness analysis. Public Health Nutr 2019 Oct 14;23(1):83-93. [doi: 10.1017/s1368980019002945]
52. Nghiem N, Leung W, Cleghorn C, Blakely T, Wilson N. Mass media promotion of a smartphone smoking cessation app: modelled health and cost-saving impacts. BMC Public Health 2019 Mar 08;19(1):283 [FREE Full text] [doi: 10.1186/s12889-019-6605-8] [Medline: 30849943]
53. van der Deen FS, Wilson N, Cleghorn CL, Kvizhinadze G, Cobiac LJ, Nghiem N, et al. Impact of five tobacco endgame strategies on future smoking prevalence, population health and health system costs: two modelling studies to inform the tobacco endgame. Tob Control 2018 May;27(3):278-286. [doi: 10.1136/tobaccocontrol-2016-053585] [Medline: 28647728]

54. Elley CR, Garrett S, Rose SB, O'Dea D, Lawton BA, Moyes SA, et al. Cost-effectiveness of exercise on prescription with telephone support among women in general practice over 2 years. Br J Sports Med 2011 Dec 16;45(15):1223-1229. [doi: 10.1136/bjsm.2010.072439] [Medline: 21081641]
55. Wang J, Sereika S, Chasens E, Ewing, Matthews, Burke L. Effect of adherence to self-monitoring of diet and physical activity on weight loss in a technology-supported behavioral intervention. Patient Prefer Adher 2012 Mar:221. [doi: 10.2147/ppa.s28889]
56. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Inform J 2016 Jun;22(2):184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]
57. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
58. Patel A, Schofield GM, Kolt GS, Keogh JW. General practitioners' views and experiences of counselling for physical activity through the New Zealand Green Prescription program. BMC Fam Pract 2011 Nov 02;12(1):119 [FREE Full text] [doi: 10.1186/1471-2296-12-119] [Medline: 22044577]
59. Annual data explorer 2019/20: New Zealand health survey [data file]. Ministry of Health. 2020. URL: https://minhealthnz.shinyapps.io/nz-health-survey-2019-20-annual-data-explorer/_w_0bdad694/#!/ [accessed 2021-06-30]
60. Goodman A, Sahlqvist S, Ogilvie D, iConnect Consortium. New walking and cycling routes and increased physical activity: one- and 2-year findings from the UK iConnect Study. Am J Public Health 2014 Sep;104(9):38-46. [doi: 10.2105/AJPH.2014.302059] [Medline: 25033133]

Abbreviations

CHD: coronary heart disease
GP: general practitioner
MET: metabolic equivalent of task
mHealth: mobile health
NZ: New Zealand
PMSLT: proportional multistate life table
PN: practice nurse
QALY: quality-adjusted life year

Edited by R Kukafka; submitted 30.06.21; peer-reviewed by V Gc, KK Lim; comments to author 22.09.21; revised version received 04.10.21; accepted 29.10.21; published 20.12.21.

Please cite as:
Grout L, Telfer K, Wilson N, Cleghorn C, Mizdrak A
Prescribing Smartphone Apps for Physical Activity Promotion in Primary Care: Modeling Study of Health Gain and Cost Savings
J Med Internet Res 2021;23(12):e31702
URL: https://www.jmir.org/2021/12/e31702
doi:10.2196/31702
PMID:34931993

Original Paper

Explaining Online Information Seeking Behaviors in People With Different Health Statuses: German Representative Cross-sectional Survey

Elena Link[1], DPhil; Eva Baumann[1], DPhil; Christoph Klimmt[1], DPhil

Department of Journalism and Communication Research, Hanover University of Music, Drama and Media, Hanover, Germany

Corresponding Author:
Elena Link, DPhil
Department of Journalism and Communication Research
Hanover University of Music, Drama and Media
Expo Plaza 12
Hanover, 30539
Germany
Phone: 49 5113100474
Email: Elena.Link@ijk.hmtm-hannover.de

Abstract

Background: Worldwide, the internet is an increasingly important channel for health information. Many theories have been applied in research on online health information seeking behaviors (HISBs), with each model integrating a different set of predictors; thus, a common understanding of the predictors of (online) HISB is still missing. Another shortcoming of the theories explaining (online) HISB is that most existing models, so far, focus on very specific health contexts such as cancer. Therefore, the assumptions of the Planned Risk Information Seeking Model (PRISM) as the latest integrative model are applied to study online HISB, because this model identifies the general cognitive and sociopsychological factors that explain health information seeking intention. We shift away from single diseases and explore cross-thematic patterns of online HISB intention and compare predictors concerning different health statuses as it can be assumed that groups of people perceiving themselves as ill or healthy will differ concerning their drivers of online HISB. Considering the specifics of online HISB and variation in individual context factors is key for the development of generalizable theories.

Objective: The objective of our study was to contribute to the development of the concept of online HISB in 2 areas. First, this study aimed to explore individual-level predictors of individuals' online HISB intention by applying the postulates of PRISM. Second, we compared relevant predictors of online HISB in groups of people with different health statuses to identify cross-thematic central patterns of online HISB.

Methods: Data from a representative sample of German internet users (n=822) served to explain online HISB intentions and influencing patterns in different groups of people. The applicability of the PRISM to online HISB intention was tested by structural equation modeling and multigroup comparison.

Results: Our results revealed PRISM to be an effective framework for explaining online HISB intention. For online HISB, attitudes toward seeking health information online provided the most important explanatory power followed by risk perceptions and affective risk responses. The multigroup comparison revealed differences both regarding the explanatory power of the model and the relevance of predictors of online HISB. The online HISB intention could be better explained for people facing a health threat, suggesting that the predictors adopted from PRISM were more suitable to explain a problem-driven type of information-seeking behavior.

Conclusions: Our findings indicate that attitudes toward seeking health information online and risk perceptions are of central importance for online HISB across different health-conditional contexts. Predictors such as self-efficacy and perceived knowledge insufficiency play a context-dependent role—they are more influential when individuals are facing health threats and the search for health information is of higher personal relevance and urgency. These findings can be understood as the first step to develop a generalized theory of online HISB.

(J Med Internet Res 2021;23(12):e25963) doi:10.2196/25963

KEYWORDS

online health information seeking behavior; Planned Risk Information Seeking Model; health status; theory building; personal survey

Introduction

Relevance of Focusing on Online Health Information Seeking Behavior

The internet occupies an increasingly important site for health information in many regions of the world [1,2]. For instance, in the United States, 80.2% of the population search for health-related information via the internet [2]. In Germany, using the internet to seek health-related information is also of increasing relevance. A national survey showed that 72% of the population use the internet to seek information on health issues [3]. The internet provides an active, problem-oriented opportunity to find a high volume and variety of online health information available virtually anywhere and anytime. That information can be used to guide individuals' health-related decisions, help to cope with uncertainties, and find strategies for living with health threats. Online health information seeking behaviors (HISBs) give individuals more control over their health care and greater knowledge about their condition [4,5]. Doing so, online HISB is a "key facilitator for promotion, maintaining, and returning to health" [6]. Research on why people seek health information online can identify which specific groups can be reached using online sources for health interventions [7]. Therefore, it is important to learn more about the online search for health information and to understand individuals' choices that determine their online HISB.

A wide range of theories has been applied or used as guidance to examine online HISB, with each model integrating a different set of predictors [5,6,8-12]. Thus, a common understanding of predictors of online HISB is still missing [13]. To learn more about the dominant predictors of online HISB, this study aims to explore individual-level influencing factors that lead people to seek health information online. Further, we adopted the Planned Risk Information Seeking Model (PRISM; [14]) as the latest integrative model for the online domain and test its applicability. Another shortcoming of the theories explaining HISB and online HISB that our study aims to address is that most existing models, so far, focus on very specific health contexts such as cancer or diabetes [9,15-18]. Contrary to this focus, a trend in information seeking research follows Case's [19] call to shift away from single diseases [14,20,21]. In response to this requirement, we aim to explore cross-thematic patterns of online HISB. We use health status as a comparative dimension because it can be assumed that groups of people perceiving themselves as ill or healthy will differ concerning the personal relevance of online HISB and its drivers [21-23]. Considering variations in individual context factors is the key for the development of generalizable theories because it helps to identify both context-specific determinants and cross-contextually important predictors that would emerge as the core of online HISB [6].

To sum up, the objective of our study was to contribute to the theoretical development of the concept of online HISB concerning 2 areas. First, we aimed to apply the PRISM to online HISB to analyze predictors of online HISB in general. Second, we compared the relevant predictors of online HISB in groups of people with different health statuses to identify cross-thematic central patterns of online HISB.

Approaches and Challenges of Health Information Seeking in the Online Domain

HISB is a complex, often multistage, process that can be conceptualized by its triggers, channel selection, search strategy, types of information sought, and outcomes [24]. Johnson and Case [6] describe the selection of a channel as the most basic decision individuals can make regarding information seeking. According to the assumptions of a goal-directed selection of information channels, the specific characteristics of a channel influence the intention to turn to it [25-29]. Even though the combined use of multiple channels for information seeking is prevalent [14,19,30], we argue that a comprehensive understanding of the different channel-specific processes of HISB is needed. Therefore, we consider internet-specific capabilities, attitudes, and norms as access points for further theory development [8,31]. Compared with other channels, the internet offers permanent, 24-hour access to extensive, multifaceted, in-depth, and latest health information. Furthermore, the active and goal-oriented options to search for health information online allow a high degree of customizability [11,12,32]. It is also suitable to seek sensitive information in contexts where anonymity is of high relevance [33]. Beyond channel characteristics, explanations for why people go online have been found in specific predictors. Powell and colleagues [34] described the characteristics of online HISB identifying certain motivations, such as the desire for reassurance, a second opinion, and better understanding of information. However, information with varying quality raises the importance of certain abilities and perceived internet self-efficacy to find reliable and accurate health information increases [12,32,35]. In sum, existing theory and research on HISB should be extended by taking channel characteristics of the internet into account and modeling online HISB.

Modeling Online HISBs

The high importance of the internet for HISB has motivated a large number of studies, most of which aim to describe and explain internet use for health-related purposes [36]. Because of the wide range of theories and predictors used to examine online HISB [8,9], it remains unclear which predictors are the dominant or universal ones for online HISB. A specific model of online HISB is missing; thus, in our study, a well-established model of HISB serves as a foundation for modeling online HISB to take past theoretical progress in general HISB into account to allow progress. The assumptions of the PRISM [14] are adopted to study online HISB because this model identifies the general cognitive and sociopsychological factors that motivate HISB. Variables considered in PRISM were already applied to

examine online HISB [13], but the whole model was not examined regarding online HISB so far.

The PRISM was developed with a general health reference and aimed at explaining information seeking intention in light of individual-level factors across different risk- and health-related issues that are valid for multichannel HISB [14]. PRISM is an integrated model postulating the importance of 7 distal and proximal factors applied from models such as the Theory of Planned Behavior [37], the Risk Information Seeking and Processing Model (RISP; [22,23]), and the Comprehensive Model of Information Seeking [8,26]. It is often viewed as an expanded iteration of the RISP that has incorporated additional models to explain general information seeking intentions as an outcome [30]. Because of this comprehensiveness, we deemed the PRISM particularly suitable for our approach.

Based on the Theory of Planned Behavior, the PRISM posits that the intention to seek information is the result of attitudes toward information seeking, subjective seeking-related norms, as well as perceived seeking control. Attitudes toward information seeking capture an individuals' evaluation of the information-seeking behavior. Subjective norms consist of perceived expectations or social pressure of one's social surroundings to seek information (injunctive norms) as well as seeking behaviors perceived in one's social surroundings (descriptive norms). The perceived behavioral control is understood as an individual's ability to seek information [14]. Further, adopted from the RISP, the PRISM includes an individual's perception of his/her state of knowledge and knowledge insufficiency as well as health-related risk perception and affective risk response [21-23] as factors that influence the intention to search for information. Knowledge and knowledge insufficiency depict the assumption that the desired level of confidence in an individuals' knowledge motivates information seeking [14,21]. Knowledge insufficiency is the gap between what an individual knows and what he or she needs to know to feel confident [23,30]. Risk perception refers to the susceptibility and severity of risks to one's health, whereas affective risk response describes a negative affective reaction induced by the health risk [14,21]. All predictors are postulated to be positively related to the HISB intention.

Kahlor [14] postulated that the predictors integrated into the PRISM are valid across channels, but the claim has never been empirically tested and the individual explanatory power of the predictors remains uncertain. Particularly, some past studies including review and meta-analysis indicated that the power of the predictors of online HISB differs from that of offline sources [8,31,38,39]. Therefore, Wang and coauthors [13] concluded that there is a need to refine the PRISM for online HISB.

Concerning attitudes toward seeking and subjective norms, a meta-analytical review [13] showed that the relation between both predictors and online HISB is understudied. About the attitude, recent research postulated a positive association between attitudes toward online HISB and online HISB intention [13], suggesting that the postulates of the PRISM apply to the online context. Comparing source usage regarding the association with norms, the first findings are mixed. Although studies support that subjective norms are a critical antecedent of online HISB [40], others found that normative influences of family and friends motivate the choice of interpersonal and mass media channels but not the internet [39]. Accordingly, the results at least call into question whether norms are among the strongest predictors of online HISB, as presumed for HISB [21].

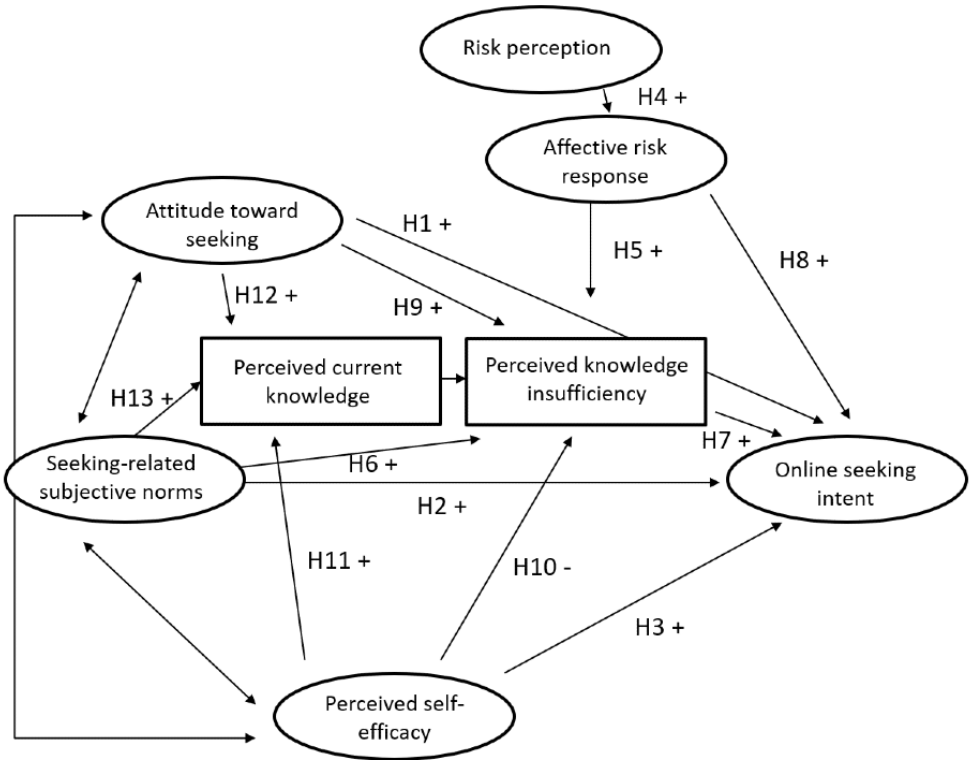
Besides, recent studies stressed that particular (perceived) abilities gain importance for health-related internet use compared with other sources [13,31]. Instead of perceived seeking control that is considered in the PRISM [14], perceived self-efficacy is often theorized as a factor influencing internet use [8,13,35]. The terms are sometimes used synonymously, or self-efficacy is understood as a dimension of seeking control [20,41]. However, although perceived seeking control includes internal abilities and external capabilities such as the accessibility of information, self-efficacy is predominantly based on internal control factors referring to beliefs in one's capabilities to execute a certain course of action such as online HISB [42]. Therefore, we categorize perceived self-efficacy as a part of perceived seeking control [43]. Because more than 90% of German residents possess access to the internet [44] and Germany is characterized by a high information and communication technology development index, external barriers to information acquisition are perceived as low, whereas own capabilities might be a relevant predictor of online HISB. In line with studies showing that self-efficacy is a well-confirmed predictor of internet use [35], we aim to focus on perceived internal control factors of those residents having access to the internet. Acquiring health information online is understood as a behavior that requires many unique capabilities from learning how to use the internet or search engines to gather health information, locating different perspectives or high-quality information, and evaluating the quality of health websites. The need for corresponding abilities underscores the importance of efficacy perceptions to perform online HISB, as self-efficacy is known to determine how much effort an individual is willing to invest to perform a certain course of action such as online HISB [35].

The multichannel comparison of Yang et al [39] highlighted that risk perception and affective risk responses as well as current knowledge and knowledge insufficiency had different impacts on HISB depending on the channel used. For online HISB, risk perception, current knowledge, and knowledge insufficiency were shown to be of low relevance for online HISB [39]. For negative affective risk responses, a weak but a significant positive association with online HISB was supported, whereas negative responses were not associated with turning to interpersonal or other media sources [39]. Given past findings that the internet often serves as a channel for a second opinion [11,35,45], prior use of other sources might result in higher perceived knowledge or higher awareness of knowledge insufficiency as predictors of online HISB and influence the risk perception and affective risk responses, which might influence the selection of the internet.

Overall, these findings indicate that established factors of HISB such as social norms, informational self-efficacy, risk perception, affective risk response, and knowledge insufficiency are of relevance for online HISB and increase the intention to seek health information online. However, the dominance of the predictors might differ in comparison to HISB in offline

channels. Therefore, it is indicated to test whether the general assumptions of the PRISM can be applied for explaining online HISB. Our first research question (RQ1) is as follows: *Can the PRISM be applied to explain online HISB intention?* The single hypotheses (H1 to H13) reflect the predicted relationships of PRISM, whereas the single predictors were transferred to online HISB. The single hypotheses are shown in Table 1 and illustrated in Figure 1.

Figure 1. Predicted relationships of the Planned Risk Information Seeking Model applied to online health information seeking behavior. H: hypothesis.



Health Status–Related Importance of Predictors of Online HISB

Aiming to explain online HISB beyond single disease contexts, we consider the personal relevance of health topics [39] distinguishing between routine (ie, driven by a general interest) and nonroutine information seeking (ie, triggered by a specific health challenge; [23,33,46]). This context factor of online HISB is considered by exploring differences in the importance of predictors of online HISB between a generally healthy population and people who are facing current health issues. Current research considering health-related context factors of online HISB mainly considered health status as a predictor of online HISB frequency but did not distinguish predictors of online HISB depending on different health-related contexts [1,4,9,10,42]. However, the PRISM and RISP have already been tested concerning different communities and different risks, that is, across differing health-related contexts [16,23,39,47-50]. Findings revealed the consistency of factor importance across different risks [23,51] but showed first indications of differences between populations with limited versus high personal relevance of a health topic [39]. More detailed findings of Yang et al [39] suggested that in the general population, higher perceived risks increased online HISB, whereas among patients with cancer, risk perception caused a decrease of online HISB. Furthermore, people's sense of knowledge insufficiency varied based on health status, and normative beliefs had a stronger impact on online HISB in the general population than that among patients with cancer [39,51,52]. Further, the attitudes toward seeking information online can be determined by health status as they are associated with the salience of perceived channel characteristics [28] and the perceived utility of the internet [8]. As perceived health threats may reduce the self-ascribed ability to succeed in online HISB [1,53], it is conceivable to posit a context-dependent role of self-efficacy to use the internet for health-related purposes. Recent research has suggested that health status may result in different influence patterns of online HISB, which might contradict the cross-channel universality of the PRISM. To substantiate this assumption, a comparison of the empirical importance of predictors in groups of individuals with different health statuses is required. Thus, our second research question (RQ2) is as follows: *Do the direct and indirect predictors of online HISB intention differ between people who face a health threat and people who perceive themselves as healthy?*

Methods

Recruitment Procedure and Participants

We conducted a standardized personal survey with a representative sample of 1001 German individuals. The fieldwork was conducted by a German market research company using quota sampling intended to guarantee representative population data. Because we aimed to explain online HISB, our analysis included only people with internet access (822/1001, 82.1% of the respondents). The respondents were aged between 18 and 88 years (mean 47.08 [SD 16.29] years), and more than half of the participants were females (459/822, 55.8%).

Regarding school education, the largest proportion of participants had a secondary school leaving certificate (371/822, 45.1%). Approximately 28.6% (235/822) of respondents had less than a secondary school leaving certificate, 14.8% (122/822) had graduated from high school, and 11.4% (94/822) had university degrees.

Measures: Online Seeking Intention

Based on Kahlor's study [14], 3 out of 5 items were applied to the internet that, on the one hand, express a different strength of intention ("I intend to look for health information on the internet in the near future;" "I will look for health information related to my personal health and risks to my health in the near future") and, on the other hand, aim at intensifying the search for information in the future ("I intend to find more information about my health on the internet soon"). A global query that refers to the internet as a whole seemed justified, as recent studies [45,54] and our survey confirmed that search engines were central (mean 3.36 [SD 0.96]; frequency measured on a 5-point Likert-type scale ranging from 1 "never" 1 to 5 "very often"), while direct access to specific online communities (mean 1.68 [SD 0.95]) or health websites (mean 2.04 [SD 1.01]) was rarely practiced. Further, in the United States, 77% of online health seekers reported using a search engine to begin a search, whereas only 13% began at a site that is specialized in health information [54]. The 3 items were rated on a 5-point Likert-type scale ranging from strongly disagree (1) to strongly agree (5) (α=.920; mean 2.45 [SD 1.13]).

Measures: Predictors

Attitudes Toward Online Information Seeking

We adopted the measurement from Kahlor [14] and instructed our participants to describe their attitudes toward online HISB with seven 5-point semantic differential pairs (eg, bad/good, unhelpful/helpful; α=.944; mean 3.29 [SD 0.90]).

Seeking-Related Social Norms

To measure injunctive and descriptive norms, we used single items adopted from Kahlor [14]: "My family and friends expect me to seek information on health-related topics and risks" (mean 2.75 [SD 1.15]; injunctive norm) and "People in my life whose opinions I value seek information on risks and health-related topics" (mean 3.37 [SD 0.98]; descriptive norm). Both items were evaluated on a 5-point Likert-type scale: strongly disagree (1) to strongly agree (5). Since injunctive and descriptive norms are theoretically conceptualized as 2 dimensions of social norms [52,55,56] and their correlation is mediocre (r=0.38; *P*=.01), we decided to treat them as distinct concepts.

Perceived Health-Related Internet Self-efficacy

Internet self-efficacy was assessed based on a scale of Eastin and LaRose [57], which was adapted and applied to online HISB by Rains [35]. The measure consists of 8 items describing different tasks (eg, evaluating the quality of health websites, finding high-quality health information) asking for respondents' perceived abilities in using the internet to acquire health information. Participants' applicability to the statements was measured on a 5-point Likert-type scale (α=.947; mean 3.51 [SD 0.97]).

Risk Perception and Affective Risk Response

Following Kahlor [14], we measured risk perception asking for the susceptibility ("How likely are you to become ill in the next year?") and the severity of an illness (eg, "If you were to become ill in the next year, how serious do you think it would be?"). All items for risk perception were measured on a Likert-type scale ranging from 1 (not at all) to 10 (extremely; α=.863; mean 3.14 [SD 1.89]). As affective responses to perceived health risks, we assessed the extent to which the participants felt worried, scared, or uncertain (α=.963; mean 3.32 [SD 2.21]). The items that capture affective risk response were measured on a Likert-type scale ranging from 1 (not at all) to 10 (extremely).

Perceived Knowledge and Knowledge Insufficiency

In line with Kahlor [14], we assessed knowledge and knowledge insufficiency on scales from 0 to 100. First, we asked the respondents to rate their current state of knowledge about health-related topics (mean 57.69 [SD 23.01]). Second, they were asked to rate their needed level of knowledge to cope with health-related topics (mean 73.49 [SD 19.7]).

Health Status

Health status was measured in 2 different ways [58]. First, respondents were asked to rate their health status (mean 3.35 [SD 0.85]) on a 5-point scale from "poor" (1) to "excellent" (5). Second, we asked whether the participants perceived themselves as currently completely healthy, acutely slightly ill, acutely seriously ill, or chronically ill [58]. Additionally, we provided the option to refuse to answer this question. To have a sufficient group size for comparison, the answers "acutely slightly ill" (154/822, 18.7%) or "acutely seriously ill" (16/822, 1.9%) and "chronically ill" (59/822, 7.2%) were merged (see RQ2). For this purpose, 2 groups—people perceiving themselves as healthy (n=564) or ill (n=229)—were built (see Multimedia Appendix 1, Table S1). Of the 822 participants, 29 (3.5%) refused to answer the question about their health status and were not considered for the group comparison (see RQ2). To validate this measurement, the relationship between both measurements of health status was assessed showing a rather strong association (r=0.53; *P*=.01).

Statistical Analysis

To answer our research questions and test our hypotheses, we used latent-variable structural equation modeling in R (version 3.4.2). We used two-step modeling to verify all measurement models before testing the structural model. The data fit of all measurement models was evaluated as satisfying (see Multimedia Appendix 1). In response to the second research question, we conducted a multigroup analysis to test the PRISM applied to online HISB intention in a group comparison for healthy and ill internet users. Therefore, measurement and structural invariance were evaluated. The results for measurement invariance are shown in Multimedia Appendix 1. The invariance appeared problematic but justifiable in single cases (see Multimedia Appendix 1). The structural invariance was determined by comparing the unconstrained and constrained model using χ^2 and fit statistics.

Results

RQ1 asked if the PRISM predictors can be applied to explain online HISB. The model fit indices showed a satisfactory model fit (χ^2_{335}=846.5, P≤.001; comparative fit index [CFI]=0.962, root mean square error of approximation [RMSEA]=0.043, standardized root mean square residual [SRMR]=0.054; [59,60]). Since the other indices had very satisfactory levels, the significant χ^2-test was attributed to the sample size [61,62]. The model accounted for 31.8% of the variance in online HISB intention. Besides, the model explained 21.9% of the variance in perceived knowledge insufficiency, 5.3% of the variance in perceived knowledge, and 68.5% of the variance in affective risk response. The results of the hypotheses tests (standardized beta coefficients and their significance) are reported in Table 1 and Figure 2. Overall, 7 of the 13 hypotheses were confirmed. Paths that could not be confirmed applied to online HISB were associated with the distinction between injunctive and descriptive norms. Only influences of injunctive norms were found. Injunctive norms were positively related to online HISB intention (see H2) and perceived knowledge (see H13) but negatively related to perceived knowledge insufficiency (see H6). Additional relationships that could not be supported for online HISB intention included the paths between affective risk response and perceived knowledge insufficiency (see H5), between perceived self-efficacy and perceived knowledge insufficiency (see H10), and between attitudes toward seeking and perceived knowledge (see H11). In contrast to previous studies focusing on HISB [14,47,50], the path between perceived knowledge insufficiency and online HISB intention was confirmed (see H7).

Table 1. Overview of the hypotheses and outcomes.

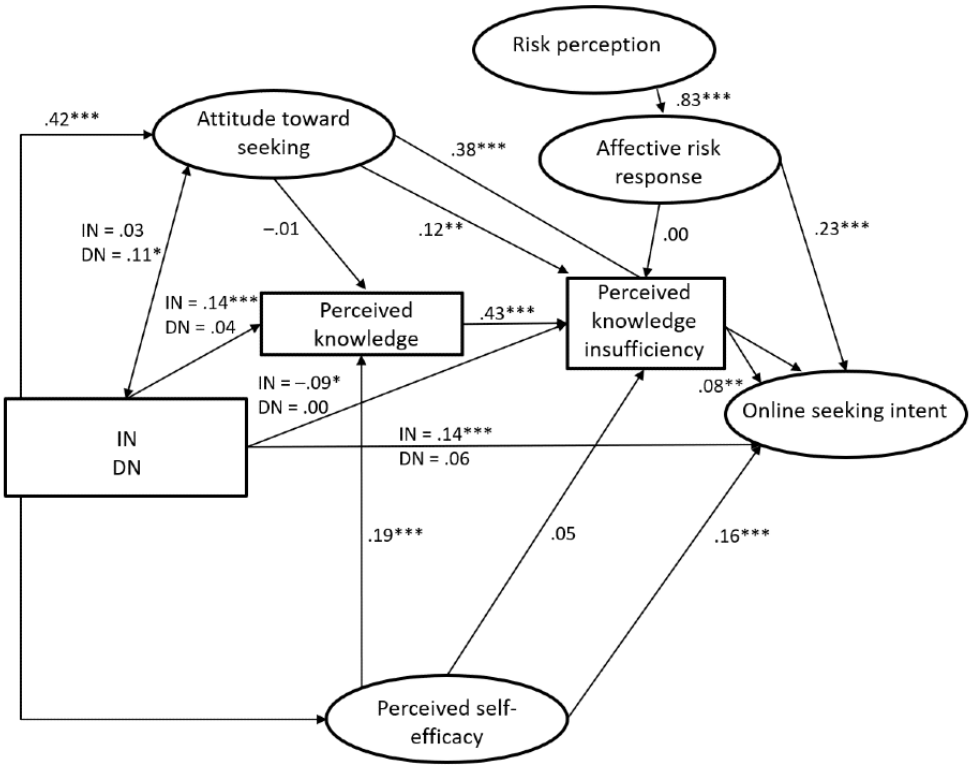
Proposed path (H: hypothesis)	Online PRISM[a]	Online PRISM (group: healthy)	Online PRISM (group: ill)
H1: Attitude toward seeking (online) will be positively related to online information–seeking intent.	Supported	Supported	Supported
H2: Seeking-related subjective/injunctive and descriptive norms will relate positively to online information–seeking intent.	IN[b]: Supported DN[c]: Not supported (not significant)	IN: Supported DN: Not supported (not significant)	IN: Not supported (not significant) DN: Supported
H3: Perceived seeking control/self-efficacy to search for information (online) will be positively related to online information–seeking intent.	Supported	Not supported (not significant)	Supported
H4: Risk perceptions will be positively related to affective risk response.	Supported	Supported	Supported
H5: Affective risk response will relate positively to perceived knowledge insufficiency.	Not supported (not significant)	Not supported (not significant)	Not supported (not significant)
H6: Seeking related subjective/injunctive and descriptive norms will relate positively to perceived knowledge insufficiency.	IN: Not supported (negatively related) DN: Not supported (not significant)	IN: Not supported (negatively related) DN: Not supported (not significant)	IN: Not supported (negatively related) DN: Not supported (not significant)
H7: Perceived knowledge insufficiency will relate positively to information seeking intent.	Supported	Not supported (not significant)	Supported
H8: Affective risk response will be positively related to (online) information–seeking intent.	Supported	Supported	Supported
H9: Attitude toward seeking (online) will relate positively to perceived knowledge insufficiency.	Supported	Not supported	Supported
H10: Perceived seeking control/self-efficacy will be negatively related to perceived knowledge insufficiency.	Not supported (not significant)	Not supported (not significant)	Not supported (not significant)
H11: Attitude toward seeking will be positively related to perceived knowledge.	Not supported (not significant)	Not supported (not significant)	Not supported (not significant)
H12: Perceived seeking control/self-efficacy will be positively related to perceived knowledge.	Supported	Supported	Supported
H13: Seeking-related subjective/injunctive and descriptive norms will be positively related to perceived knowledge.	IN: Supported DN: Not supported (not significant)	IN: Supported DN: Not supported (not significant)	IN: not supported (not significant) DN: Not supported (not significant)

[a]PRISM: Planned Risk Information Seeking Model.

[b]IN: injunctive norm.

[c]DN: descriptive norm.

Figure 2. Planned Risk Information Seeking Model applied to online health information seeking behavior intention. The results of the hypotheses tests are shown as standardized beta coefficients and their significance. DN: descriptive norm; IN: injunctive norm. *P<.05, **P<.01, ***P<.001.



RQ2 asked for differences of predictors of online HISB between groups of people perceiving themselves as healthy or ill. Comparing the unconstrained and constrained model showed that both models fit the data fairly well (unconstrained: χ^2_{686}=1295.0, P≤.001; CFI=0.958; RMSEA=0.050; SRMR=0.061; constrained: χ^2_{726}=1295.0, P≤.001; CFI=0.949; RMSEA=0.051; SRMR=0.065), but the χ^2 difference test indicated that the models are not equivalent ($\Delta\chi^2$=101.1, Δdf=40; P≤.001). The unconstrained model was superior, implying that path coefficients vary among groups. The comparison of the power of the model for healthy and ill people showed that the model accounted for a higher level of variance for ill people with regard to online HISB intention (healthy people: R^2=0.273 vs ill people: R^2=0.437; ΔR^2=0.164), the affective response to perceived risks (healthy people: R^2=0.550 vs ill people: R^2=0.693; ΔR^2=0.143), and perceived risks (healthy people: R^2=0.058 vs ill people: R^2=0.074; ΔR^2=0.016). In turn, the model explains more variance in perceived knowledge insufficiency for healthy people (R^2=0.229) in comparison with ill people (R^2=0.193; ΔR^2=0.036).

Looking at single paths (see Figure 3 and Figure 4), we found different influences from social norms (see H2 and H13). While in the healthy group, injunctive norms had weak to mediocre but significant effects on seeking intention (healthy: β=.18; P<.001; ill: β=.08; P=.20), on perceived level of knowledge (healthy: β=.19; P<.001; ill: β=.01; P=.91), and on perceived knowledge insufficiency (healthy: β=–.12; P=.01; ill: β=–.04; P=.55), these paths were not significant or weaker in the group of ill people (Figure 4).

For the group of ill people, the intention to seek information was significantly influenced by descriptive norms (healthy: β=.02; P=.61; ill: β=.13; P=.047) rather than injunctive norms. Furthermore, perceived internet self-efficacy had a significant and stronger influence on the online HISB intention (see H3; β=.09; P=.07 vs ill: β=.30; P<.001) and the perceived level of knowledge among ill people (see H12; healthy: β=.11; P=.049 vs ill: β=.28; P<.001). The association between self-efficacy and attitudes toward seeking was stronger in the group of healthy people compared with people facing health threats (healthy: β=.47; P<.001; ill: β=.30; P<.001). Other differences related to the relationships between perceived knowledge insufficiency and health information seeking intention (see H7; healthy: β=.05; P=.19; ill: β=.17; P=.01) and between the attitudes toward seeking information online and knowledge insufficiency (see H9; healthy: β=.05; P=.31; ill: β=.18; P=.01). In both cases, only the path for the group of people perceiving themselves as ill was significant.

Figure 3. Online Planned Risk Information Seeking Model comparing healthy and ill people (group of healthy individuals). The results of the tests are shown as standardized beta coefficients and their significance. DN: descriptive norm; IN: inductive norm. *P<.05, **P<.01, ***P<.001.

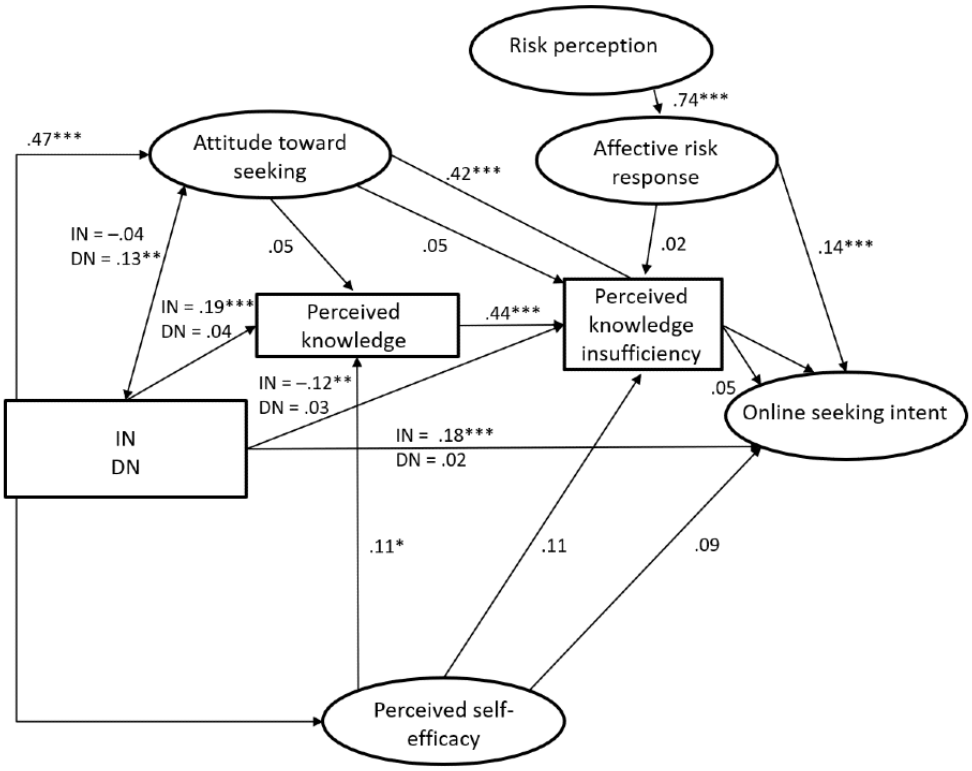
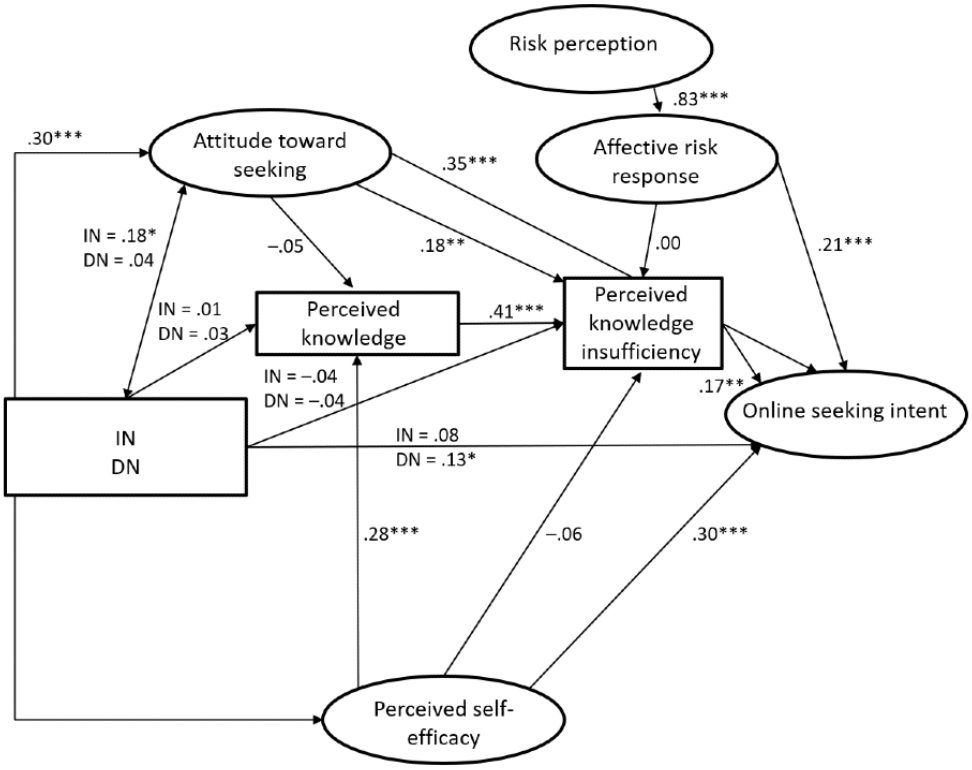


Figure 4. Online Planned Risk Information Seeking Model comparing healthy and ill people (group of ill individuals). The results of the tests are shown as standardized beta coefficients and their significance. DN: descriptive norm; IN: inductive norm. *P<.05, **P<.01, ***P<.001.



Discussion

Specifics of the Predictors of Online HISB Intention

As the internet is a frequently used channel for health-related purposes, we applied the PRISM [14] to the internet to provide a theoretically sound analysis of predictors of online HISB intention and refine the PRISM for the online context (RQ1). Our results reveal PRISM to be an effective framework for explaining online HISB intention. For online HISB, attitudes toward seeking health information online provide the most important explanatory power followed by risk perceptions and

affective risk responses. These factors are identified as most essential for online HISB intention and can be understood as the cross-contextual core or dominant predictors of online HISB [6,13]. Further, perceived health-related internet self-efficacy, injunctive norms, and perceived knowledge insufficiency influence a higher online HISB intention. However, they seem to be of secondary importance to explain the intention to turn to the internet. For self-efficacy, our findings do not confirm the postulated higher importance for online HISB [31]. This may be attributed to the fact that our respondents perceive internet use for health-related purposes as less challenging than assumed. As the analyzed sample consists of internet users only, it would be a valuable supplement to compare users with nonusers. However, this was not the aim of this study but should be considered in the future. Regarding the rather weak impact of social norms, our findings are in line with the results of Yang et al [39]. Social norms seem to be less important for online HISB in comparison with HISB in general [14,16]. This is particularly evident for descriptive norms, which have neither an effect on the intention to seek information nor on perceived knowledge insufficiency and the current state of knowledge. The rather weak effect of perceived knowledge insufficiency should be assessed against the background of previous studies that found no relation between knowledge insufficiency and seeking intention [14,50]. This raises the question about the importance of perceived knowledge as a relevant predictor of seeking intention [21]. Our results indicate that the PRISM is a fruitful base to explain online HISB intentions and demonstrate that the process of online HISB rests partially on influencing factors other than offline HISB [21,39]. Thus, more research comparing online HISB with HISB in other channels is needed to determine the differences between channels and to develop a robust generalizable theory of online HISB.

Impact of Health Status on Online HISB

RQ2 addressed the difference between predictors determining online HISB intentions for groups of people perceiving themselves as healthy or ill. The multigroup comparison revealed differences both regarding the explanatory power of the model and the relevance of different influencing factors of online HISB. In general, the causal assumptions were confirmed with greater effect sizes in the subsample of ill people than that in the healthy subsample. Accordingly, the online HISB intention can be better explained for people facing a health threat, suggesting that the influencing factors adopted from PRISM are more suitable to explain a problem-driven type of information-seeking behavior [33,46,63]. These findings call for improving the theoretical understanding of online HISB intentions by healthy people who do not feel immediate pressure to acquire health-related knowledge even though they might perceive their knowledge as insufficient. At the same time, these results indicate that the internet is a particularly important source for ill people [3,45,64], which is important to consider in efforts of distributing health information.

Referring to the single predictors, attitudes toward seeking and risk perception are identified as important in both groups. Although attitudes toward seeking have a stronger influence in the group of healthy people, risk perception and affective risk response show a stronger association for people facing health threats. This contrasts with the findings of Yang et al [39], who identified a positive effect of risk perception on internet use for healthy people and a negative effect for patients with cancer. The severity of a disease could be responsible for this difference. It should be pointed out that the group of people facing a health threat surveyed in this study covers very different health statuses; hence, it is impossible to address differences related to diseases with different degrees of severity and susceptibility. Specific influencing patterns are plausible, and non-HISB coping strategies such as information avoidance are more likely in the case of highly severe diseases such as cancer [65].

Beyond the cross-contextual factors of online HISB, social norms, internet self-efficacy, and perceived knowledge insufficiency are observed as context-specific influencing factors of online HISB intentions. The results regarding the role of social norms are in line with the findings of Yang et al [39] and complement them. They confirm that injunctive norms are a more important predictor for healthy than for ill people. Besides, our data show a higher relevance of descriptive norms for people perceiving health threats. Thus, the individual adherence to injunctive and descriptive norms indicates that healthy people are more strongly influenced for online HISB by perceived expectations held by people in their social environment, while higher intentions among people perceiving themselves as ill are more strongly influenced by the perceived behavior of relevant others. This can be traced back to the fact that affected individuals observe others as role models in a challenging situation; adopting information-seeking behaviors they perceive others to apply in similar situations may provide a solution for the individual search for help, while the perceived pressure from others appears less important in this situation. The context-dependent role of self-efficacy shows that online HISB intention is only directly affected by self-efficacy for people facing health threats that might result from a higher pressure to search for health-related information in the case of illness. For healthy people, the perception of being capable to gather information appears less relevant as their search for health information lacks urgency. However, self-efficacy has a stronger effect on positive attitudes toward seeking information online. This might indicate that healthy individuals assume that they can benefit from the search for information if and when the corresponding needs arise [66]. Another difference between individuals in contexts of health versus illness was found regarding the influence of perceived knowledge insufficiency on the intention for online HISB. This influence is only significant in the group of ill people, which indicates that sufficient knowledge is particularly perceived relevant when problems occur and information might help; so far, this path has not been confirmed in many studies [14,50]. Besides, the influence of attitudes toward seeking health information on perceived knowledge insufficiency was only apparent for the group of people facing health threats. This suggests that health threats are raising awareness of knowledge insufficiency [39]. Overall, the results suggest that the relative importance of predictors of online HISB differs depending on the individual's health status. Therefore, considering health status is a valuable extension of theory-based explanatory modeling of online HISB.

Limitations and Resulting Tasks for Future Research

Although this study informs about the necessity to adjust theoretical models for online HISB, the limitations of the study need to be considered. First, it should be mentioned that information seeking rarely involves only one channel [19,28,29]. Future studies should therefore map the influences of other sources on selecting the internet for health information and compare individual selection factors for different channels. A second limitation is that our study is based on cross-sectional data; therefore, no causality statements can be derived. To ensure a deeper understanding of the processes of information search, longitudinal designs are required in future research. Furthermore, there is a need for cross-cultural studies because the use of the internet to search for health-related topics varies between cultures and countries. We already know that people in the United States are overall more likely to use the internet for health purposes than Europeans [1,2,67]. This suggests that HISB should be systematically compared in terms of cultural context. Third, it should be critically reflected that the categorization of respondents as healthy or ill is unspecific and only based on self-perceptions. Owing to the comparatively small number of cases, differences between acute and chronic illness could not be considered, and different degrees of severity and courses of diseases could not be mapped separately. It can be assumed that, depending on different diseases and different indicators of being or feeling more or less healthy or ill, intentions of HISB will vary [39,66,68]. It should also be noted that the health status variable used to compare the model between contexts of health and illness is closely linked to risk perception as a component of the model. The state of health represents the actual state and can be understood as a predictor of risk perception directed toward the future. Instead of comparing groups of people with different health statuses, further research should integrate health status as a predictor influencing risk perception. Another limitation is associated with the not sufficiently complex measurement of intention to seek information online, as there are manifold types of message forms and contents that can be accessed online (eg, dedicated websites, social media, streaming services, video platforms with user-generated content). Future research should thus consider the diversity to use the internet. Likewise, injunctive and descriptive norms were measured as single items, which should be improved in follow-up studies as well.

Conceptual Perspectives: Advancing Theorizing on HISB

To sum up, our results indicate that attitudes toward seeking health information online and risk perceptions are of central importance for online HISB across different health statuses. The importance of social norms is generally low for online HISB. Further, predictors such as self-efficacy and perceived knowledge insufficiency play a context-dependent role—they are more influential when individuals are facing health threats and the search for health information is of higher personal relevance and urgency. These findings can be understood as the first step to develop a generalized theory of online HISB. In general, the findings from invariance tests indicate that some of the applied measurements such as risk perception may have limited equivalence in health and illness contexts. This points to limited generalizability of the PRISM, which seems to reach greater explanatory power for ill people who are facing more immediate information needs than for healthy people whose inclination to seek and acquire health-related information is not that much energized by situational circumstances [33,46]. This conceptual challenge also holds implications for measurement and testing the PRISM. Although items should be used that are independent of specific health status, the common way of measuring appears not to be optimal for application across different groups with different health statuses. Thus, further studies are needed to show the extent to which the differences found are valid for different health statuses and different information sources to advance model development and increase understanding of the processes of HISB in more detail.

Application Perspectives

Effective disseminating of health information in the online domain can benefit from these findings, as the obtained importance of predictors in health and illness conditions allows to characterize target audiences with greater precision. One striking example of such possible benefits is the observation that internet-related self-efficacy makes a substantially greater difference for online HISB among the ill than among the healthy respondents. Providers of information on a specific illness should thus be aware that online platforms may fail to reach out to relevant parts of the target group who are affected by the illness, as only those with higher self-efficacy are likely to access the online information. Additional (channel) strategies beyond online services are thus advised to avoid information underserving of those patients who do not believe in their ability to find and acquire online health information. In contrast, providers of general or prevention-related health information should consider that motivation of healthy individuals to access their content online is primarily influenced by a positive attitude. Thus, although internet resources have low access barriers and are (seemingly) easy to find and acquire, population segments who do not find online HISB desirable are unlikely to make use of the offered information. Hence, among healthy people, the availability, importance, and usefulness of preventive health information on the internet need to be precisely triggered to increase information pull; besides, general health information should still be transported through mass communication channels such as billboards or television that reach out to nonsearching audiences even in times of nearly permanent internet availability and use.

Acknowledgments

Our work has been carried out as an integral part of the Biofabrication for Niedersächsisches Zentrum für Biomedizintechnik, Implantatforschung und Entwicklung (NIFE) Initiative, which is financially supported by the ministry of Lower Saxony and the VolkswagenStiftung. NIFE is the Lower Saxony Center for Biomedical Engineering, Implant Research and Development, a joint

translational research center of the Hannover Medical School, the Leibniz University Hannover, the University of Veterinary Medicine Hannover, and the Laser Center Hannover.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 31 KB - jmir_v23i12e25963_app1.docx]

References

1. Bachl M. Online health information seeking in Europe: Do digital divides persist? SC|M 2016;5(4):427-453. [doi: 10.5771/2192-4007-2016-4-427]
2. Health information national trends survey. National Cancer Institute. 2017. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?PK_Cycle=9&qid=757 [accessed 2021-11-30]
3. Link E, Baumann E. Use of health information on the internet: personal and motivational influencing factors. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Jun;63(6):681-689 [FREE Full text] [doi: 10.1007/s00103-020-03144-5] [Medline: 32367207]
4. Bjarnadottir RI, Millery M, Fleck E, Bakken S. Correlates of online health information-seeking behaviors in a low-income Hispanic community. Inform Health Soc Care 2016 Dec;41(4):341-349 [FREE Full text] [doi: 10.3109/17538157.2015.1064429] [Medline: 26837012]
5. Hartoonian N, Ormseth SR, Hanson ER, Bantum EO, Owen JE. Information-seeking in cancer survivors: application of the Comprehensive Model of Information Seeking to HINTS 2007 data. J Health Commun 2014;19(11):1308-1325. [doi: 10.1080/10810730.2013.872730] [Medline: 24742287]
6. Johnson J, Case DO. Health Information Seeking. New York, NY: Peter Lang Inc, International Academic Publishers; First printing edition; 2012:1433118246.
7. Kreps GL. Strategic use of communication to market cancer prevention and control to vulnerable populations. Health Mark Q 2008;25(1-2):204-216. [doi: 10.1080/07359680802126327] [Medline: 18935885]
8. Basnyat I, Nekmat E, Jiang S, Lin J. Applying the Modified Comprehensive Model of Information Seeking to Online Health Information Seeking in the Context of India. J Health Commun 2018;23(6):563-572. [doi: 10.1080/10810730.2018.1493058] [Medline: 29979921]
9. Van Stee SK, Yang Q. Online Cancer Information Seeking: Applying and Extending the Comprehensive Model of Information Seeking. Health Commun 2018 Dec;33(12):1583-1592. [doi: 10.1080/10410236.2017.1384350] [Medline: 29083231]
10. Rains SA. Health information seeking and the World Wide Web: an uncertainty management perspective. J Health Commun 2014;19(11):1296-1307. [doi: 10.1080/10810730.2013.872731] [Medline: 24730566]
11. Lin J, Dutta MJ. A Replication of Channel Complementarity Theory among Internet Users in India. Health Commun 2017 Apr;32(4):483-492. [doi: 10.1080/10410236.2016.1140268] [Medline: 27301884]
12. Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What Predicts Online Health Information-Seeking Behavior Among Egyptian Adults? A Cross-Sectional Study. J Med Internet Res 2017 Jun 22;19(6):e216 [FREE Full text] [doi: 10.2196/jmir.6855] [Medline: 28642216]
13. Wang X, Shi J, Kong H. Online Health Information Seeking: A Review and Meta-Analysis. Health Commun 2021 Sep;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
14. Kahlor L. PRISM: a planned risk information seeking model. Health Commun 2010 Jun;25(4):345-356. [doi: 10.1080/10410231003775172] [Medline: 20512716]
15. Higgins O, Sixsmith J, Barry M, Domegan C. A literature review on health information-seeking behavior on the web: A health consumer and health professional perspective. European Centre for Disease Prevention and Control (ECDC). URL: https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Literature%20review%20on%20health%20information-seeking%20behaviour%20on%20the%20web.pdf [accessed 2021-11-30]
16. Hovick SR, Kahlor L, Liang M. Personal cancer knowledge and information seeking through PRISM: the planned risk information seeking model. J Health Commun 2014;19(4):511-527 [FREE Full text] [doi: 10.1080/10810730.2013.821556] [Medline: 24433251]
17. Ybarra M, Suman M. Reasons, assessments and actions taken: sex and age differences in uses of Internet health information. Health Educ Res 2008;23(3):512-521. [doi: 10.1093/her/cyl062] [Medline: 16880222]
18. Hovick SR, Liang M, Kahlor L. Predicting cancer risk knowledge and information seeking: the role of social and cognitive factors. Health Commun 2014;29(7):656-668. [doi: 10.1080/10410236.2012.763204] [Medline: 24093914]
19. Case DO. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Amsterdam, Boston: Emerald Group Publishing; 3rd edition; 2007.

20. Kahlor LA. An Augmented Risk Information Seeking Model: The Case of Global Warming. Media Psychology 2007;10(3):414-435. [doi: 10.1080/15213260701532971]
21. Yang ZJ, Aloe AM, Feeley TH. Risk Information Seeking and Processing Model: A Meta-Analysis. J Commun 2014;64(1):20-41. [doi: 10.1111/jcom.12071]
22. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
23. Dunwoody S, Griffin R. Risk information seeking and processing model. In: The SAGE Handbook of Risk Communication. Thousand Oaks: Sage Publications; 2015:102.
24. Galarce E, Ramanadhan S, Viswanath K. Health information seeking. In: The Routledge Handbook of Health Communication. 2nd Edition. New York, NY: Routledge; 2011.
25. Blumler J, Katz E. The Uses of Mass Communications: Current Perspectives on Gratifications Research. Beverly Hills, California: Sage Publications; 1974.
26. Johnson J, Meischke H. A Comprehensive Model of Cancer-Related Information Seeking Applied to Magazines. Human Comm Res 1993;19(3):343-367. [doi: 10.1111/j.1468-2958.1993.tb00305.x]
27. Knobloch-Westerwick S. Choice and Preference in Media Use: Advances in Selective Exposure Theory and Research. New York: Routledge; 2015.
28. Rains SA, Ruppel EK. Channel Complementarity Theory and the Health Information-Seeking Process. Communication Research 2013;43(2):232-252. [doi: 10.1177/0093650213510939]
29. Ruppel EK, Rains SA. Information Sources and the Health Information-Seeking Process: An Application and Extension of Channel Complementarity Theory. Communication Monographs 2012;79(3):385-405. [doi: 10.1080/03637751.2012.697627]
30. Yang ZJ, Kahlor L, Li H. A United States-China Comparison of Risk Information–Seeking Intentions. Communication Research 2013;41(7):935-960. [doi: 10.1177/0093650213479795]
31. Percheski C, Hargittai E. Health information-seeking in the digital age. J Am Coll Health 2011;59(5):379-386. [doi: 10.1080/07448481.2010.513406] [Medline: 21500056]
32. Dobransky K, Hargittai E. Inquiring minds acquiring wellness: uses of online and offline sources for health information. Health Commun 2012;27(4):331-343. [doi: 10.1080/10410236.2011.585451] [Medline: 21932982]
33. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
34. 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;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
35. Rains SA. Seeking Health Information in the Information Age: The Role of Internet Self-Efficacy. Western Journal of Communication 2008;72(1):1-18. [doi: 10.1080/10570310701827612]
36. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
37. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
38. Carcioppolo N, Yang F, Yang Q. Reducing, Maintaining, or Escalating Uncertainty? The Development and Validation of Four Uncertainty Preference Scales Related to Cancer Information Seeking and Avoidance. J Health Commun 2016;21(9):979-988. [doi: 10.1080/10810730.2016.1184357] [Medline: 27450905]
39. Yang ZJ, McComas KA, Gay G, Leonard JP, Dannenberg AJ, Dillon H. Information Seeking Related to Clinical Trial Enrollment. Communication Research 2010;38(6):856-882. [doi: 10.1177/0093650210380411]
40. Veinot TC, Meadowbrooke CC, Loveluck J, Hickok A, Bauermeister JA. How "community" matters for how people interact with information: mixed methods study of young men who have sex with other men. J Med Internet Res 2013;15(2):e33 [FREE Full text] [doi: 10.2196/jmir.2370] [Medline: 23428825]
41. Kahlor L, Rosenthal S. If We Seek, Do We Learn? Science Communication 2009;30(3):380-414. [doi: 10.1177/1075547008328798]
42. Bandura A. Self-Efficacy: The Exercise of Control. New York, NY: Worth Pub; 1st edition; 1997.
43. Armitage CJ, Conner M. Social cognition models and health behaviour: A structured review. Psychology & Health 2000 Mar;15(2):173-189. [doi: 10.1080/08870440008400299]
44. Frees B, Koch W. ARD/ZDF-onlinestudie 2018: Zuwachs bei medialer internetnutzung und kommunikation. Media Perspektiven. 2018. URL: https://www.ard-zdf-onlinestudie.de/files/2018/0918_Frees_Koch.pdf [accessed 2021-12-03]
45. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
46. van der Rijt GAJ. Health Complaints and the Search for Health Information. Communications 2000;25(2):143-160. [doi: 10.1515/comm.2000.25.2.143]

47. Hubner AY, Hovick SR. Understanding Risk Information Seeking and Processing during an Infectious Disease Outbreak: The Case of Zika Virus. Risk Anal 2020;40(6):1212-1225. [doi: 10.1111/risa.13456] [Medline: 32067244]
48. Griffin RJ, Zheng Yang, ter Huurne E, Boerner F, Ortiz S, Dunwoody S. After the Flood. Science Communication 2008;29(3):285-315. [doi: 10.1177/1075547007312309]
49. ter Huurne EFJ, Griffin RJ, Gutteling JM. Risk Information Seeking Among U.S. and Dutch Residents. Science Communication 2009;31(2):215-237. [doi: 10.1177/1075547009332653]
50. Willoughby JF, Myrick JG. Does Context Matter? Examining PRISM as a Guiding Framework for Context-Specific Health Risk Information Seeking Among Young Adults. J Health Commun 2016;21(6):696-704. [doi: 10.1080/10810730.2016.1153764] [Medline: 27187057]
51. Kahlor L, Dunwoody S, Griffin RJ, Neuwirth K. Seeking and Processing Information about Impersonal Risk. Science Communication 2016;28(2):163-194. [doi: 10.1177/1075547006293916]
52. Rimal RN, Real K. Perceived Risk and Efficacy Beliefs as Motivators of Change. Human Comm Res 2003;29(3):370-399. [doi: 10.1111/j.1468-2958.2003.tb00844.x]
53. Jiang S, Street RL. Pathway Linking Internet Health Information Seeking to Better Health: A Moderated Mediation Study. Health Commun 2017;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
54. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2018-03-16]
55. Cialdini RB, Reno RR, Kallgren CA. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology 1990;58(6):1015-1026. [doi: 10.1037/0022-3514.58.6.1015]
56. Rimal RN, Lapinski MK, Cook RJ, Real K. Moving toward a theory of normative influences: how perceived benefits and similarity moderate the impact of descriptive norms on behaviors. J Health Commun 2005;10(5):433-450. [doi: 10.1080/10810730591009880] [Medline: 16199387]
57. Eastin M, LaRose R. Internet Self-Efficacy and the Psychology of the Digital Divide. Journal of Computer-Mediated Communication 2000;6(1). [doi: 10.1111/j.1083-6101.2000.tb00110.x]
58. Goldner M. How health status impacts the types of information consumers seek online. Information, Communication & Society 2006;9(6):693-713. [doi: 10.1080/13691180601063982]
59. 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;6(1):1-55. [doi: 10.1080/10705519909540118]
60. Urban D, Mayerl J. Strukturgleichungsmodellierung. Wiesbaden: Springer Fachmedien; 2014.
61. Bearden WO, Sharma S, Teel JE. Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models. Journal of Marketing Research 1982;19(4):425. [doi: 10.2307/3151716]
62. Hoelter JW. The Analysis of Covariance Structures. Sociological Methods & Research 1983;11(3):325-344. [doi: 10.1177/0049124183011003003]
63. Griffin RJ, Neuwirth K, Dunwoody S, Giese J. Information Sufficiency and Risk Communication. Media Psychology 2004;6(1):23-61. [doi: 10.1207/s1532785xmep0601_2]
64. Zschorlich B, Gechter D, Janßen IM, Swinehart T, Wiegard B, Koch K. Health information on the Internet: Who is searching for what, when and how? Z Evid Fortbild Qual Gesundhwes 2015;109(2):144-152. [doi: 10.1016/j.zefq.2015.03.003] [Medline: 26028452]
65. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Libr Assoc 2005;93(3):353-362 [FREE Full text] [Medline: 16059425]
66. Oh YS, Song NK. Investigating Relationships Between Health-Related Problems and Online Health Information Seeking. Comput Inform Nurs 2017;35(1):29-35. [doi: 10.1097/CIN.0000000000000234] [Medline: 26950091]
67. Alvarez-Galvez J, Salinas-Perez JA, Montagni I, Salvador-Carulla L. The persistence of digital divides in the use of health information: a comparative study in 28 European countries. Int J Public Health 2020;65(3):325-333. [doi: 10.1007/s00038-020-01363-w] [Medline: 32277246]
68. Oh YS, Cho Y. Examining the relationships between resources and online health information seeking among patients with chronic diseases and healthy people. Soc Work Health Care 2015;54(2):83-100. [doi: 10.1080/00981389.2014.987940] [Medline: 25674723]

Abbreviations

CFI: comparative fit index
HISB: health information seeking behavior
PRISM: Planned Risk Information Seeking Model
RISP: Risk Information Seeking and Processing Model
RMSEA: root mean square error of approximation
RQ: research question

SRMR: standardized root mean square residual

Edited by R Kukafka, G Eysenbach; submitted 23.11.20; peer-reviewed by S Van Stee, K Hung, S Kanthawala; comments to author 22.02.21; revised version received 19.04.21; accepted 15.10.21; published 10.12.21.

Please cite as:
Link E, Baumann E, Klimmt C
Explaining Online Information Seeking Behaviors in People With Different Health Statuses: German Representative Cross-sectional Survey
J Med Internet Res 2021;23(12):e25963
URL: https://www.jmir.org/2021/12/e25963
doi:10.2196/25963
PMID:34890348

Original Paper

The Mediating Influence of the Unified Theory of Acceptance and Use of Technology on the Relationship Between Internal Health Locus of Control and Mobile Health Adoption: Cross-sectional Study

Ashraf Sadat Ahadzadeh[1], PhD; Shin Ling Wu[2], PhD; Fon Sim Ong[3], PhD; Ruolan Deng[4], MA

[1]Department of Journalism, Xiamen University Malaysia, Selangor, Malaysia

[2]Department of Psychology, School of Medical and Life Sciences, Sunway University, Kuala Lumpur, Malaysia

[3]Malaysian Research Institute on Ageing, Universiti Putra Malaysia, Selangor, Malaysia

[4]Department of Communication, University of Vienna, Vienna, Austria

Corresponding Author:
Ashraf Sadat Ahadzadeh, PhD
Department of Journalism
Xiamen University Malaysia
Bandar Sunsuria, Sepang
Selangor, 43900
Malaysia
Phone: 60 147393423
Email: ahadzadeh1980@gmail.com

Abstract

Background: Mobile health (mHealth) as an innovative form of information and communications technology can efficiently deliver high-quality health care by enhancing communication and health management, reducing costs, and increasing access to health services. An individual's internal health locus of control (HLOC) is found to be associated with the behavioral intent to adopt mHealth. However, little is known about the underlying mechanism of this association.

Objective: The primary objective of this study was to test the mediation influence of the Unified Theory of Acceptance and Use of Technology (UTAUT) on the relationship between internal HLOC and the behavioral intention to use mHealth.

Methods: A total of 374 responses were collected from Malaysian adult users of mHealth, using convenience and snowball sampling methods. Partial least squares structural equation modeling was used to analyze the data. Data were collected for variables, including demographics, internal HLOC, and modified UTAUT constructs (ie, performance expectancy, effort expectancy, and social influence).

Results: The results showed that there was no direct relationship between internal HLOC and the behavioral intention to use mHealth (β=−0.039, *P*=.32). The indirect relationship between internal HLOC and the intent to adopt mHealth was supported, indicating that the UTAUT constructs performance expectancy (β=0.104, *P*<.001), effort expectancy (β=0.056, *P*=.02), and social influence (β=0.057, *P*=.002) mediated this relationship. The results showed full mediation, with total variance explained at 47.2%.

Conclusions: This study developed an integrative model, where a health-related disposition (internal HLOC), mHealth-related beliefs (performance expectancy and effort expectancy), and normative pressure (social influence) were combined to explain the underlying mechanism of the behavioral intent to adopt mHealth. The results showed that the intention to adopt mHealth is mediated by the influence of UTAUT factors, while HLOC has no direct effect on adoption intention. The findings provide insights into augmenting mHealth adoption among the public by enhancing the perceived benefits of mHealth, helping design more effective and user-friendly mHealth tools, and capitalizing on social normative influence to adopt mHealth. This study utilized the constructs of the UTAUT model to determine the intention to use mHealth. Future research should focus on other health- and technology-related theories to ascertain other possible factors influencing the behavioral intent of mHealth adoption.

(J Med Internet Res 2021;23(12):e28086) doi:10.2196/28086

KEYWORDS

mobile health; mHealth; internal health locus of control; performance expectancy; effort expectancy; social influence; mediation

Introduction

Background

Over the past decades, health care systems in most countries around the world have experienced tremendous changes because of the rapid advancement in information and communications technology (ICT). Mobile health (mHealth), as an innovative form of ICT, is one of the most prominent services with remarkable effects on the development of the health care system [1]. According to the Global Observatory for eHealth, mHealth is defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [2]. mHealth has the potential to enhance the quality of health care systems by improving communication, efficiency, treatment adherence, and health/disease management; reducing costs; and increasing access to interventions and health services [3-6].

The popularity of mHealth programs has grown worldwide as evidenced by a Statista report in 2017 regarding the estimated number of mHealth app downloads, which has exponentially increased from 1.7 billion in 2013 to 3.7 billion in 2017 [7]. A global survey undertaken by the World Health Organization in 2011 showed that 114 countries have established mHealth initiatives, with health call centers, emergency toll-free telephone services, management of emergencies and disasters, and mobile telemedicine being the 4 most frequently reported initiatives in many countries [8]. Malaysia in Southeast Asia has started restructuring its health care policies and encouraging more start-ups to use disruptive technology to solve key medical challenges [9]. As Malaysia has a high smartphone usage (78% in 2018 [10]), mHealth adoption in the country can be a new and effective approach to empower people in health management and transform the attitude toward health care from reactive to proactive [11].

The importance and implications of mHealth have inspired researchers to investigate the factors in the adoption of mHealth. A cluster of studies viewed mHealth as a perceived technology–driven behavior and attempted to find the correlates of such behavior using technology adoption theories [12-14]. Another stream of studies attempted to explain mHealth from health-related perspectives and examined health factors as predictors of mHealth adoption [15]. While the former has been widely researched, the latter has received less attention.

Among health-related factors, the belief that health events are caused by one's own actions is one of the major predictors of health behaviors such as greater engagement in health/disease management, healthier lifestyle, and better physical and mental quality of life [16,17]. Those who believe that they have an active role in one's own health, also known as internal health locus of control (HLOC), are more likely to take responsibility toward their health and engage in health behaviors [18]. Given that individuals' health-related dispositional factors are significant antecedents of health behaviors [19], including the use of health technologies [20-22], very limited studies have examined the relationship between internal HLOC and the intent to adopt mHealth [23].

Although the direct relationship between internal HLOC and behavioral intention to use mHealth provides a noteworthy tenet to knowledge, little is known regarding the underlying mechanism of this relationship. Focusing on mediating factors, which facilitate the relationship between internal HLOC and the intent to use mHealth, can provide a better insight to Malaysian health policy makers and health care professionals for embedding mHealth use in daily life and promoting mHealth functions (such as health/disease information seeking, communicating for health-related purposes, and downloading and using any health-related apps) for health management and more importantly active participation in disease prevention, thus reducing the need for health care services and consequently the toll on the Malaysian health care system.

Therefore, this study aims to contribute to the literature by introducing 3 constructs of the Unified Theory of Acceptance and Use of Technology (UTAUT) (ie, performance expectancy, effort expectancy, and social influence) as mediators between internal HLOC and the intention to use mHealth, which, to the best of our knowledge, has not been examined thus far. The UTAUT is one of the most widely accepted technology adoption theories with a wide applicability and a high explanatory power to predict the intent to adopt technology [24]. By integrating the UTAUT constructs as mediators in the association between internal HLOC and the intention to use mHealth, a mediation model has been developed to explain the mechanism that underlies the relationship between internal HLOC and behavioral intent to adopt mHealth. Therefore, the objectives of this study are to investigate (1) the relationship between internal HLOC and behavioral intent to adopt mHealth; (2) the relationship between the constructs of the UTAUT (ie, performance expectancy, effort expectancy, and social influence) and the intention to adopt mHealth; (3) the relationship between internal HLOC and the constructs of the UTAUT; and (4) the mediation effects of performance expectancy, effort expectancy, and social influence in the relationship between internal HLOC and the intent to adopt mHealth.

Literature

HLOC and Technology-Related Behaviors

Locus of control (LOC) is a psychological construct that is derived from the social learning theory of personality [25]. Wallston et al [18] developed a multidimensional HLOC scale, which denotes how much individuals believe they are in control of their current and future health. HLOC can be measured as an internal or external HLOC [18]. Internal HLOC posits an active role in one's own health and taking responsibility toward health, whereas external HLOC is divided into 2 parts, with one referring to powerful others and the other referring to chance, luck, or the influence of religion [18].

In general, higher levels of internal HLOC are more likely to drive healthy behaviors and more successful changes in health behaviors and preventive health behaviors, whereas higher levels of external HLOC are not. Stronger internal HLOC orientations were found to be related to greater engagement in health-enhancing behaviors (such as exercise and diet) [16,26,27]; better mental and physical quality of life [16]; lower rates of smoking and excessive drinking [28,29]; better

medication adherence [30]; lower levels of stress, depression, and anxiety [16,31]; longer survival time after surgery [32]; and better health rehabilitation and care [33,34]. Moreover, individuals with internal HLOC tend to actively use coping strategies that focus on solving problems [17] and show greater beliefs in the ability to control the risk of disease [35,36]. Conversely, those with high chance HLOC believe that there is nothing much they can do to influence their health outcomes, and consequently, they are less likely to engage in health behaviors [37]. Additionally, those with higher external LOC are more likely to report higher levels of stress [38] and have worse mental health, as they use more emotion-focused strategies [17].

Research into technology adoption has relied on LOC as a construct that explains adoption behavior. Empirical studies provided support for the association between LOC and higher propensity of adopting technology in an array of technologies where differences in internal and external LOC tend to differentiate the behaviors between these 2 groups [39-42]. Individuals with internal LOC tend to use technology in a goal-directed manner; they are more likely to adopt problem-solving stances to change the environment compared with externals [43]. Internals are more likely to have higher utilization of the technology [44,45]. They are early adopters of technology, are more satisfied with their skills to use technology, and are more comfortable with technology [43], and they perceive difficulties in using technology as associated with their own lack of abilities [46]. Given the rising technology use across all demographic groups, an emerging cluster of scholarly works has been devoted to internal HLOC to predict the intention to adopt mHealth [23].

The Mediating Effect of UTAUT Constructs

Limited research on the relationship between internal HLOC and behavioral intent to adopt mHealth calls for further investigation into the possibility of other variables that could underlie this relationship. Therefore, this study intends to suggest the mediating effect of UTAUT constructs to test the indirect relationship between internal HLOC and the intention to use mHealth. Venkatesh et al developed a unified model that has an overall comprehensive explanatory power to conceptualize and predict acceptance behavior, known as the UTAUT [47]. Numerous UTAUT studies have verified that 3 core constructs (ie, performance expectancy, effort expectancy, and social influence) can affect the intention to adopt technology [48-54]. Performance expectancy refers to “the degree to which an individual believes that using the system will help him or her to attain gains in job performance” [47]. Effort expectancy is defined “as the degree of ease associated with the use of the system” [47]. Social influence is defined “as the degree to which an individual perceives that important others believe he or she should use the new system” [47]. Performance expectancy, effort expectancy, and social influence have direct effects on behavioral intention, which in turn predicts use behavior [47].

Since its emergence, the UTAUT has been empirically tested across domains [53-55], including eHealth and mHealth [56,57]. Past studies have shown that performance expectancy, effort expectancy, and social influence are significantly associated with the intention to use eHealth and mHealth in elderly people and citizens dealing with a health problem [12,52]. Application of the UTAUT to examine the intention to use eHealth and mHealth among clinicians and health care professionals has also been endorsed in a handful of studies [14,56]. It is understood that the target population in most past studies was mainly older adults, people with a health problem, and health professionals. This study attempts to investigate the associations between the 3 constructs of UTAUT (ie, performance expectancy, effort expectancy and social influence, and intention to use mHealth) among Malaysians aged 18 years or above. Evidence showed that although most Malaysians have limited knowledge about mHealth, they reported having a positive attitude toward mHealth [58]. Likewise, a favorable affective feeling was also reported toward telemonitoring systems by patients [59], telemedicine technology by health providers [60], e-counseling by counselors [61], and internet usage for health-related purposes (such as health information seeking and communication) by women [20].

Individuals who score high in internal LOC, also known as internals, cherish innovative ideas [62], tend to be early adaptors of innovative products [44,63], and perceive technology as useful [41,42,46], thus showing a stronger tendency toward technology adoption [64]. They also master their learning environment using more proactive approaches and believe that using technology is free of effort [65]. Internals are characterized by high self-efficacy, which is the impetus to overcome difficulties in using technologies [55]. They are more likely to demonstrate a positive attitude toward computers [65,66], and greater confidence and lower levels of anxiety in using computers [67]. Internal LOC has also been empirically found to be associated with social influence [68,69]. Individuals with higher internal LOC tend to be less likely to be influenced by others [70].

In light of the above literature, we would expect internal HLOC to predict UTAUT constructs, which may, in turn, predict behavioral intent to adopt mHealth. In other words, instead of a direct relationship between HLOC and mHealth, we would expect an indirect relationship that could provide an underlying mechanism to explain how those high on internal HLOC are disposed toward mHealth use. Individuals who tend to assign the cause of health outcomes to their internal characteristics rather than to outside forces are more likely to perceive that mHealth is useful and easy to use. However, they may not use mHealth because of social influences as they do not believe that external forces, such as others, can motivate them to use mHealth. Because of their belief in mHealth usefulness and ease of use, they may have an intention to adopt mHealth. However, their resistant to social influence may hinder them from adopting mHealth. Hypotheses and justifications for the hypotheses are presented in Textbox 1.

Textbox 1. Hypotheses and justifications for the hypotheses.

Hypothesis 1: Internal health locus of control (HLOC) has a positive relationship with the intention to adopt mobile health (mHealth).

Justifications for the hypothesis

- It was found that there is a relationship between locus of control (LOC) and technology adoption in developing agriculture [39], online games [40], and online learning [41,42].
- Individuals with oropharyngeal head and neck cancer with a high propensity for an internal HLOC orientation showed their support toward telepractice models of care telerehabilitation [22].
- A cross-sectional study revealed that the amount of control college students believed they had over their health predicted willingness to use health apps and online health trackers [23].

Hypothesis 2: Performance expectancy has a positive relationship with the intention to adopt mHealth.

Hypothesis 3: Effort expectancy has a positive relationship with the intention to adopt mHealth.

Hypothesis 4: Social influence has a positive relationship with the intention to adopt mHealth.

Justifications for the hypotheses

- In a study to examine the intention of elderly people aged 57 to 77 years to use eHealth apps, expected performance and effort were highly related to the intention to use eHealth while social influence was not [52].
- A study revealed that Unified Theory of Acceptance and Use of Technology factors, namely effort expectancy, expectancy performance, and social influence, were significant determinants of the intention for mHealth adoption behavior in citizens like diabetic patients who were taking traditional health care services repeatedly from any medical hospital for diabetes, blood pressure, and cholesterol monitoring in the United States, Canada, and Bangladesh [12].
- A study on the intention to use a mobile electronic health record (MEHR) system in a sample of health care professionals (doctors and nurses) showed that the intention to use the MEHR system was indirectly influenced by effort expectancy and performance expectancy through attitudes toward the MEHR system, while social influence was found to be directly associated with the intention to utilize the MEHR system [14].
- Venugopal et al [56] examined clinical staff's perspectives on the usage of telemedicine and electronic health records in hospitals and found that performance expectancy, effort expectancy, and social influence have a significant impact on intention, which in turn impacts the usage behavior of electronic health records and telemedicine.

Hypothesis 5: Internal HLOC has a positive relationship with performance expectancy.

Justifications for the hypothesis

- Individuals with high internal LOC are more likely to seek new information when the information is personally relevant, and obtain valuable knowledge and skills to enhance their performance [65,71].
- Internals commonly display a favorable attitude toward technology [66].

Hypothesis 6: Internal HLOC has a positive relationship with effort expectancy.

Justifications for the hypothesis

- Individuals with internal LOC attributed perceived difficulty toward technology to their own abilities and attempted to use technology more effectively [46].
- Internals have more experience in using technologies and find technology, such as e-learning, easy to use [41,42,72,73].

Hypothesis 7: Internal HLOC has a negative relationship with social influence.

Justifications for the hypothesis

- Individuals with higher internal LOC are resistant to social influence as they feel they have more self-control and self-reinforcement over their life and things that happened to them [71].
- They are not easily persuaded and do not conform to others' influence [68].

Hypothesis 8: Performance expectancy mediates the positive relationship between internal HLOC and the intention to adopt mHealth.

Hypothesis 9: Effort expectancy mediates the positive relationship between internal HLOC and the intention to adopt mHealth.

Hypothesis 10: Social influence mediates the positive relationship between internal HLOC and the intention to adopt mHealth.

Justifications for the hypotheses

- Performance expectancy and effort expectancy were found to be positive and significant mediators among website design, customer service, and customer's intention to adopt internet banking [74].
- Performance expectancy and effort expectancy were found to be linked to user adoption in context awareness and Alipay, a third-party mobile and online payment platform [75].

- Fong et al [55] explored the mediating role of effort expectancy and perceived risk in the relationship between internal LOC and the intention to reuse mobile apps for making hotel reservations.

Methods

Participant Profiles

Among 374 participants in this study, there were 145 males and 229 females. The participant age ranged from 18 to 68 years (mean 28.01 years, SD 11.10). Almost 45% (166/374, 44.4%) of the participants were Chinese, and 40.7% (152/374) were Malays. In terms of health status, 47.4% (177/374) of the participants perceived their health status as good, 27.5% (103/374) perceived it as fair, and 18.2% (68/374) perceived it as very good. Participants were also asked whether they had an ongoing or a serious health problem that included heart disease, arthritis, or a mental health condition requiring frequent medical care, such as regular visits to doctors or daily medications. The majority (315/374, 84.3%) of the participants indicated that they did not have any ongoing or serious health problem, while 12.0% (45/374) reported that they did not know of any serious health problems. A small percentage (14/374, 3.7%) of participants had an ongoing disease or serious health problem. Lastly, regarding mobile phone usage experience, 40.4% (151/374) of the participants had 8 to 10 years of experience, 39.0% (146/374) had 4 to 7 years of experience, 15.0% (56/374) had more than 10 years of experience, and 5.6% (21/374) had 1 to 3 years of experience. Table 1 shows the demographic profile of the respondents.

Table 1. Demographic profile of the respondents.

Background variable	Value (N=374), n (%)
Gender	
Male	145 (38.8)
Female	229 (61.2)
Ethnicity	
Malay	152 (40.7)
Chinese	166 (44.4)
Indian	47 (12.5)
Others	9 (2.4)
Perceived health status	
Do not know	3 (0.8)
Poor	9 (2.4)
Fair	103 (27.5)
Good	177 (47.4)
Very good	68 (18.2)
Excellent	14 (3.7)
Disease	
Yes	14 (3.7)
No	315 (84.3)
Do not know	45 (12.0)
Mobile phone usage experience	
1-3 years	21 (5.6)
4-7 years	146 (39.0)
8-10 years	151 (40.4)
>10 years	56 (15.0)

Research Design and Procedure

This study used a questionnaire-based cross-sectional design to collect the required data. A total of 400 questionnaires were distributed to Malaysian adults residing in Kuala Lumpur, Malaysia. The subjects for this study were drawn from mHealth users. A screening self-report question was included in the survey to identify mHealth users. Participants were asked if they have ever used their smartphones for any health-related purposes, such as seeking health- and disease-related information online, texting messages for health-related purposes (such as reminders/alerts for appointments, taking medications,

and consultations), and downloading and using any health-related apps (such as fitness apps, and apps for health tracking and medication tracking). Participants who reported having used their smartphones at least for one of these purposes were included in the analysis.

The questionnaire comprised an informed consent form, demographic profiles, and questions related to internal HLOC and the modified UTAUT constructs for mHealth use. Data were collected using convenience and snowball sampling methods. A research assistant was recruited for data collection. Participation was voluntary where confidentiality was ensured, and respondent consent was obtained before commencing the survey. Participants were given an absolute right of withdrawal at any time and without giving any reason. The protocol of the study (including the research procedure, the rights and safety of the participants, and the method of data collection) was approved by the review board of Xiamen University Malaysia to ensure compliance with research ethics. The approval number is REC-1911.01.

The following 2 rules of thumb are used for choosing the sample size when partial least square is to be used for model analysis: (1) "10 times the scale with the largest number of formative (ie, causal) indicators (note that scales for constructs designated with reflective indicators can be ignored)," and (2) "10 times the largest number of structural paths directed at a particular construct in the structural model." [76]. This study did not have any formative indicator. Therefore, the first rule of thumb was not applicable for this study. According to the former rule of thumb, the minimum sample size for this study is 70, as the largest number of structural paths in the research model is 7, which is related to the behavioral intent to adopt mHealth. Further, recent developments suggest that researchers should determine sample size through power analysis [77,78]. Power analysis determines the minimum sample size by taking into account the part of a model with the largest number of predictors [79]. Therefore, we used G*Power to determine the required sample size. Based on G*Power, the adequacy of sample size was computed using the statistical tests recommended. For a mediation model, the input information was as follows: test family: "F tests," statistical test: "linear multiple regression: fixed model, R2 deviation from zero," type of power analysis: "a priori: compute required sample size–given α, power, and effect size," with effect size=0.15, α=.05, power=0.90, and number of predictors=4. G*Power showed that the minimum sample size required for the mediation model is 108, with an actual power of 0.90. Therefore, the number of responses collected for this study was sufficient for analysis.

Measurements

Internal HLOC

In measuring internal HLOC, 6 items were adopted from the Multidimensional Health Locus of Control scale developed by Wallston et al [18]. Respondents were asked to rate items on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

Performance Expectancy, Effort Expectancy, Social Influence, and Intention to Use mHealth

Items to measure performance expectancy, effort expectancy, social influence, and intention to use mHealth were directly extracted from the original UTAUT model [24,47] and modified to make them consistent with mHealth use behavior. Participants were requested to rate the items on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Questions related to internal HLOC and the modified UTAUT constructs for mHealth use are included in Multimedia Appendix 1.

Results

Data Analysis

In this study, partial least squares structural equation modeling (PLS-SEM) was used to examine the proposed conceptual framework using SmartPLS software. By using PLS-SEM, the direct and indirect effects of multiple independent and dependent variables can be tested simultaneously, which provides greater statistical power. PLS-SEM is also able to accommodate a study with a small sample size despite the complexity of the models [79]. After excluding incomplete questionnaires, 374 responses were included in the analysis. The results consisted of a 2-stage analytical procedure, which involved a measurement model and structural model, to validate the model and test the hypotheses developed for the purpose of this study.

Measurement Model

The first step in the analysis concerning the measurement model was to examine the factor loading. In this study, the factor loadings of the items varied from 0.632 to 0.945 (Table 2). All items were retained as the factor loadings were above the recommended value of 0.6 [80]. Second, composite reliability, which measures the internal consistency of the constructs, was used to measure the reliability of the variables. According to Hair et al [79], the minimum value of composite reliability is 0.7. Table 2 shows that all the constructs yielded good composite reliability ranging from 0.873 to 0.954. Third, convergent validity was assessed using average variance extracted (AVE). Hair et al [79] mentioned that the minimum acceptable value for AVE is 0.5 or higher, which indicated that 50% or more of the items were explained by the construct. In this study, all the constructs exceeded the acceptable value, which indicated that all the constructs obtained good convergent validity (Table 2).

Table 2. Assessment results of the measurement model.

Constructs and items	Loading	CR[a]	AVE[b]
Internal health locus of control		0.873	0.535
IHLC1	0.714		
IHLC2	0.662		
IHLC3	0.799		
IHLC4	0.747		
IHLC5	0.817		
IHLC6	0.632		
Performance expectancy		0.915	0.783
PE1	0.853		
PE2	0.920		
PE3	0.880		
Effort expectancy		0.937	0.789
EE1	0.875		
EE2	0.890		
EE3	0.926		
EE4	0.861		
Social influence		0.954	0.874
SI1	0.919		
SI2	0.945		
SI3	0.940		
Behavioral intention		0.891	0.732
BI1	0.785		
BI2	0.898		
BI3	0.880		

[a]CR: composite reliability.

[b]AVE: average variance extracted.

Lastly, discriminant validity was determined using the heterotrait-monotrait (HTMT) ratio of correlation as recommended by Henseler et al [81] because it is more suitable for discriminant validity assessment compared with the Fornell-Larcker criterion and cross-loading assessment. In order to achieve discriminant validity, Henseler et al suggested that a threshold value of 0.90 or below is required [81]. As shown in Table 3, all the study variables met the criterion to establish discriminant validity where the HTMT values were below 0.90.

Table 3. Discriminant validity using the heterotrait-monotrait (HTMT) ratio.

Construct	Internal health locus of control	Performance expectancy	Effort expectancy	Social influence	Behavioral intention
Internal health locus of control	N/A[a]	0.390	0.373	0.212	0.213
Performance expectancy	0.390	N/A	0.886	0.687	0.745
Effort expectancy	0.373	0.886	N/A	0.636	0.678
Social influence	0.212	0.687	0.636	N/A	0.682
Behavioral intention	0.213	0.745	0.678	0.682	N/A

[a]N/A: not applicable.

Structural Model

Multicollinearity is assessed using the variance inflation factor (VIF). In this study, all VIF values were below 5, which indicated no violation of the multicollinearity assumption. The structural model was assessed using R^2, beta, and *t* value, which were obtained via a 5000 resampling of the bootstrapping procedure. The direct, total indirect, and specific indirect effects are shown in Table 4.

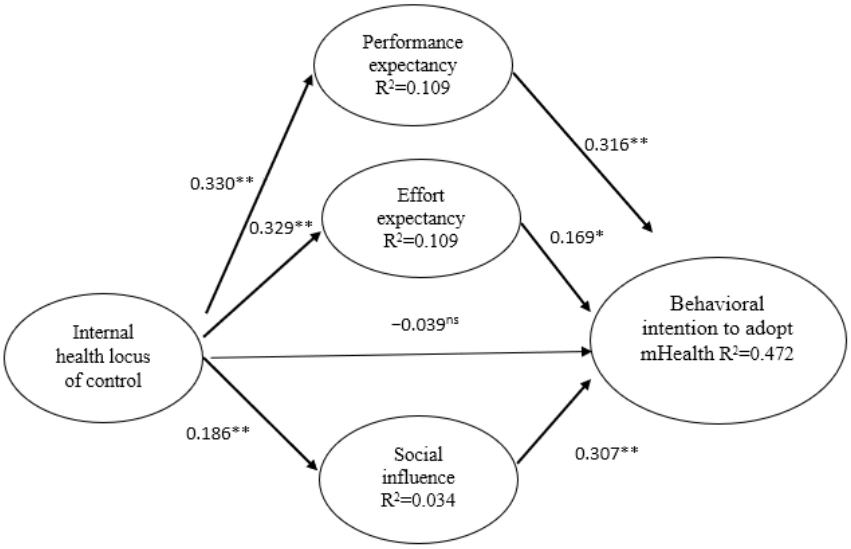
Table 4. Direct, total indirect, and specific indirect effects.

Path	R^2	Beta	*t* value	*P* value
Direct effects				
Behavioral intention	0.472			
Internal health locus of control → behavioral intention		−0.039	0.999	.32
Performance expectancy → behavioral intention		0.316	4.859	<.001
Effort expectancy → behavioral intention		0.169	2.672	.008
Social influence → behavioral intention		0.307	5.715	<.001
Performance expectancy	0.109			
Internal health locus of control → performance expectancy		0.330	6.522	<.001
Effort expectancy	0.109			
Internal health locus of control → effort expectancy		0.329	7.020	<.001
Social influence	0.034			
Internal health locus of control → social influence		0.186	3.621	<.001
Total indirect effects				
Behavior	0.472			
Internal health locus of control → behavior through performance expectancy, effort expectancy, and social influence		0.217	5.554	<.001
Specific indirect effects				
Internal health locus of control → performance expectancy → behavioral intention		0.104	3.813	<.001
Internal health locus of control → effort expectancy → behavioral intention		0.056	2.389	.02
Internal health locus of control → social influence → behavioral intention		0.057	3.123	.002

Based on the structural model, hypothesis 1 was not supported because internal HLOC (β=−0.039, *P*=.32) was not significantly related to the intention to use mHealth. Performance expectancy (β=0.316, *P*<.001), effort expectancy (β=0.169, *P*=.008), and social influence (β=0.307, *P*<.001) had a significant and positive relationship with the intention to use mHealth, which supported hypotheses 2, 3, and 4 (Figure 1). Hypotheses 5, 6, and 7 were also supported where results showed that internal HLOC had a significant positive relationship with performance expectancy (β=0.330, *P*<.001), effort expectancy (β=0.329, *P*<.001), and social influence (β=0.186, *P*<.001).

Overall, the results supported the model of this study where the total indirect effect was significant (β=0.217, *P*<.001). All the study variables explained 47.2% of the variance in the intention to use mHealth. Three specific indirect paths were generated to test hypotheses 8 to 10. Hypothesis 8 was developed to test the mediating effect of performance expectancy on the relationship between internal HLOC and the intention to adopt mHealth. The results showed that the direct effect of internal HLOC on performance expectancy and the direct effect of performance expectancy on the intention to use mHealth were significant, while the direct effect of internal HLOC on the intention to use mHealth was not significant. This indicated that performance expectancy fully mediated the relationship between internal HLOC and the intention to use mHealth with a significant indirect effect (β=0.104, *P*<.001), supporting hypothesis 8 (Figure 1).

Figure 1. Path coefficients of the structural research model. mHealth: mobile health; ns: not significant. *P<.01, **P<.001.



For hypothesis 9, effort expectancy as the mediator for the relationship between internal HLOC and the intention to adopt mHealth was tested. The direct effect from internal HLOC to effort expectancy and the direct effect from effort expectancy to the intention to use mHealth were significant, but the direct effect from internal HLOC to the intention of adopting mHealth was not significant. This indicated that effort expectancy fully mediated the relationship between internal HLOC and the intention to use mHealth (β=0.056, P=.02). Thus, hypothesis 9 was supported (Figure 1).

In Figure 1, the results showed support for hypothesis 10, where social influence significantly mediated the relationship between internal HLOC and the intention to use mHealth. The significant indirect effect (β=0.057, P=.002) revealed that internal HLOC impacted social influence, and social influence in turn affected the intention to use mHealth. Based on the findings of the mediation tests, it can be seen that the mediating effect of performance expectancy on the relationship between internal HLOC and the intention to use mHealth was the strongest, followed by social influence and effort expectancy.

Discussion

Previous studies have demonstrated the role of LOC in the tendency to use technology [39-42,55]. Since research on mHealth behavior is limited and little is known about how LOC could influence the intention to adopt mHealth technology, this study aimed to find out the factors that could explain users' behavioral intent in mHealth technology. Drawing upon 3 constructs of the UTAUT, this study attempted to provide insights on how internal HLOC delineates the intention to adopt mHealth.

In this study, internal HLOC was not found to be significantly related to the intention to use mHealth (hypothesis 1). This result is inconsistent with previous research providing evidence that individuals with internal LOC beliefs tend to utilize health apps [23]. Some studies showed a negligible relationship between HLOC and health behaviors [22,82,83], thus leading to inconsistent findings about the significance of HLOC in driving health behaviors. Calnan [84] suggested that health behaviors may not be associated with beliefs regarding control of health but rather with concerns over risky health behaviors. The absence of a relationship between internal HLOC and behavioral intent to adopt mHealth could also be attributed to potential confounding factors, such as health literacy, that could influence HLOC, which in turn affects the intention to use mHealth. Health literacy was found to be an effective factor in predicting internal orientations [85]. Individuals with higher levels of health literacy are more likely to report higher scores of internal HLOC than those with lower levels of health literacy because their capacities to obtain, process, and understand health information and services in order to make appropriate health decisions are positively associated with their belief that they have control over their health and health behaviors [86]. Another possible confounding factor that may affect internal HLOC is the participants' economic status. Research has found that individuals with lower socioeconomic status engaged in fewer health-promoting behaviors and had different expectations about their ability to influence their own health [87]. People living in economically deprived circumstances may, as a result of their experiences, also learn that they have less control over their own lives and health status [88].

The UTAUT constructs, namely performance expectancy, effort expectancy, and social influence, were found to be significant predictors, with a positive relationship for the intention to use mHealth (hypotheses 2, 3, and 4), lending support to past studies that consistently showed the association between UTAUT determinants and eHealth and mHealth adoption [12,52,56,57]. All these findings unfailingly corroborate the UTAUT model, where if users perceive mHealth as useful, easy to use, and accepted as well as suggested by important others, they will be more likely to adopt the technology. Among these 3 constructs, performance expectancy and effort expectancy have equal contributory impact on mHealth adoption, followed by social influence. These results are consistent with the findings in the study by Tavares and Oliveira [57]. They found that performance expectancy and effort expectancy predict the same variance in the use of online services such as electronic health

record portals. It suggests that usefulness and ease of use are equally central for users when evaluating mHealth services [57]. Besides, the power of important others should not be neglected in technology adoption behavior. Research has found that social influence is an essential source of motivation to utilize hospital electronic information management systems [56]. In contrast, normative pressure was found to be an insignificant force to patient portal use behavior among the elderly [89]. These contradicting results can lead us to an inference that social influence is more likely to be a driving force for technology adoption behavior among young individuals rather than older people because "older people do not follow the bandwagon effect" [90].

To further explore whether internal HLOC can be suitably applied in the UTAUT model, the relationships between internal HLOC and 3 constructs of the UTAUT (ie, performance expectancy, effort expectancy, and social influence) in the mHealth context were postulated in this study (hypotheses 5, 6, and 7). The results showed that internal HLOC had a significant positive effect on performance expectancy, effort expectancy, and social influence in mHealth use, which suggests that the more internal the users are, the higher the perceived usefulness and ease of use they will have and the more likely they will conform to important others. These results are in line with the findings of previous studies that showed the significant influence of LOC on perceived usefulness and ease of use for mobile learning adoption [40,41]. In addition, Fong et al [55] found that internals are characterized by high self-efficacy, a trait that helps individuals to overcome difficulties. Hence, they perceive higher ease of use for new technology products compared with externals [72]. As opposed to the negative association hypothesized between internal HLOC and social influence, this study found a positive relationship between these 2 constructs. This could be due to the individual differences in tendencies to conform for informational reasons but not for normative reasons [73].

In testing the mediating role of performance expectancy, effort expectancy, and social influence, this study found that these 3 constructs fully mediated the relationship between internal HLOC and the intention to use mHealth, supporting hypotheses 8, 9, and 10. Internal HLOC was positively related to the intention to adopt mHealth through performance expectancy. Effort expectancy had a mediating effect on the relationship between internal HLOC and the intention to use mHealth (hypothesis 9), similar to the results found in previous research [55]. Individuals with internal HLOC are more likely to overcome difficulties in using new technology products. They prefer to adopt mHealth because it is easy to use. Moreover, social influence was found to significantly mediate the relationship between internal HLOC and the intention to use mHealth (hypothesis 10). Internals use mHealth partly because of the perception that important others may suggest to employ mHealth. Individuals' beliefs in their own efforts and abilities to control their health drive perceptions toward mHealth, which in turn contribute to mHealth adoption. UTAUT dimensions are central to our understanding of the association between internal HLOC and the intention to adopt mHealth. The mediation results of this study are consistent with the findings in the study by Ahadzadeh et al [20]. The authors found the centrality of perceived usefulness and perceived ease of use in the relationship between health factors and internet use for health-related purposes. A strong mediating effect of health app use efficacy was also identified in the effect of health information orientation and eHealth literacy on health app use [91].

The findings of this study have several implications. Theoretically, this study has contributed to mHealth literature by investigating the direct and indirect relationships between internal HLOC and the intention to use mHealth. The indirect relationship provided a more multifaceted understanding of mHealth adoption behavior, where both health and technology come into play in the adoption decision process. Moreover, the results attested the robustness of the UTAUT in mHealth adoption. To the best of our knowledge, this study is the first attempt to examine the indirect effect of internal HLOC on the behavioral intent of mHealth. The results of the effect of UTAUT dimensions on the intention to adopt mHealth have significant implications for health providers seeking methods to enhance mHealth engagement behavior. They can leverage cognitive and normative factors related to technology (ie, performance expectancy, effort expectancy, and social influence) to increase individuals' preferences to use mHealth for health purposes.

The limitations of this study can be attributed to the urban sample concentrated in the most developed part of Malaysia. A more representative sample should be considered in future studies. Nonprobability sampling methods (ie, convenience and snowball) employed for the sample selection and the unbalanced gender makeup of the sample can jeopardize the generalizability of the results. The cross-sectional design used in this research does not provide definite information about cause and effect relationships. Moreover, social desirability bias can be a problem with self-report measurements used in this study. The framework proposed in this study predicted 42.7% of the variance in mHealth, while a more extended model encompassing more cognitive factors can augment the prediction power of the model. Given the absence of a significant association between internal HLOC and the intent to use mHealth in this research, replication studies are suggested, which can include a broader framework of multiple health-related factors (such as perceived health susceptibility, perceived health severity, perceived health status, and health consciousness) and personality factors in order to advance the frontier of knowledge regarding HLOC and technology adoption, since technology will become even more important in the future with "Industrial Revolution 4.0." Moreover, incorporating perceived health risk factors along with HLOC will enable researchers to determine whether mHealth is a proactive/preventive health behavior, a reactive behavior, or both.

Acknowledgments

This study was funded by Xiamen University Malaysia Research Grant (XMUMRF/2019/C3/IART/0004).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions related to internal health locus of control and the modified Unified Theory of Acceptance and Use of Technology constructs.
[DOCX File , 51 KB - jmir_v23i12e28086_app1.docx]

References

1. Blaya JA, Fraser HSF, Holt B. E-health technologies show promise in developing countries. Health Aff (Millwood) 2010 Feb;29(2):244-251 [FREE Full text] [doi: 10.1377/hlthaff.2009.0894] [Medline: 20348068]
2. Monitoring and Evaluating Digital Health Intervention: A Practical Guide to Conducting Research and Assessment. World Health Organization. 2016. URL: https://www.who.int/reproductivehealth/publications/mhealth/digital-health-interventions/en/ [accessed 2020-04-12]
3. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, 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]
4. Leichman ES, Gould RA, Williamson AA, Walters RM, Mindell JA. Effectiveness of an mHealth Intervention for Infant Sleep Disturbances. Behav Ther 2020 Jul;51(4):548-558 [FREE Full text] [doi: 10.1016/j.beth.2019.12.011] [Medline: 32586429]
5. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
6. Omboni S, Caserini M, Coronetti C. Telemedicine and M-Health in Hypertension Management: Technologies, Applications and Clinical Evidence. High Blood Press Cardiovasc Prev 2016 Sep;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
7. Number of mHealth app downloads worldwide from 2013 to 2017. Statista. 2019. URL: https://www.statista.com/statistics/625034/mobile-health-app-downloads/ [accessed 2020-04-18]
8. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2019-11-10]
9. Hussein Z, Oon SW, Fikry A. Consumer Attitude: Does It Influencing the Intention to Use mHealth? Procedia Computer Science 2017;105:340-344 [FREE Full text] [doi: 10.1016/j.procs.2017.01.231]
10. Internet users survey 2018. Malaysian Communication and Multimedia Commission. 2018. URL: https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/Internet-Users-Survey-2018.pdf [accessed 2019-11-12]
11. Malaysia's Telemedicine Blueprint: Leading Healthcare into the Information Age. Malaysia Ministry of Health. 1997. URL: https://www.moh.gov.my/moh/resources/auto%20download%20images/5ca1b20928065.pdf [accessed 2019-11-10]
12. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V. A generalised adoption model for services: A cross-country comparison of mobile health (m-health). Government Information Quarterly 2016 Jan;33(1):174-187 [FREE Full text] [doi: 10.1016/j.giq.2015.06.003]
13. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: A patient's perspective. Inform Health Soc Care 2017 Jan;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
14. Kim S, Lee K, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016 Jan 30;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
15. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating m-Health Acceptance from a Protection Motivation Theory Perspective: Gender and Age Differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]
16. Cheng C, Cheung MW, Lo BCY. Relationship of health locus of control with specific health behaviours and global health appraisal: a meta-analysis and effects of moderators. Health Psychol Rev 2016 Dec;10(4):460-477 [FREE Full text] [doi: 10.1080/17437199.2016.1219672] [Medline: 27556686]
17. Wilski M, Brola W, Tomczak M. Health locus of control and mental health in patients with multiple sclerosis: Mediating effect of coping strategies. Res Nurs Health 2019 Aug;42(4):296-305 [FREE Full text] [doi: 10.1002/nur.21955] [Medline: 31173383]

18. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Educ Monogr 1978;6(2):160-170. [doi: 10.1177/109019817800600107] [Medline: 689890]
19. Covey J. The role of dispositional factors in moderating message framing effects. Health Psychol 2014 Jan;33(1):52-65. [doi: 10.1037/a0029305] [Medline: 22924446]
20. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
21. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
22. Cartmill B, Wall LR, Ward EC, Hill AJ, Porceddu SV. Computer Literacy and Health Locus of Control as Determinants for Readiness and Acceptability of Telepractice in a Head and Neck Cancer Population. Int J Telerehabil 2016;8(2):49-60 [FREE Full text] [doi: 10.5195/ijt.2016.6203] [Medline: 28775801]
23. Bennett BL, Goldstein CM, Gathright EC, Hughes JW, Latner JD. Internal health locus of control predicts willingness to track health behaviors online and with smartphone applications. Psychol Health Med 2017 Dec;22(10):1224-1229 [FREE Full text] [doi: 10.1080/13548506.2017.1317354] [Medline: 28415852]
24. 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 [FREE Full text] [doi: 10.1287/isre.11.4.342.11872]
25. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 1966;80(1):1-28. [Medline: 5340840]
26. Chen S, Acton G, Shao J. Relationships among nutritional self-efficacy, health locus of control and nutritional status in older Taiwanese adults. J Clin Nurs 2010 Aug;19(15-16):2117-2127 [FREE Full text] [doi: 10.1111/j.1365-2702.2009.03186.x] [Medline: 20659192]
27. Steptoe A, Wardle J. Locus of control and health behaviour revisited: a multivariate analysis of young adults from 18 countries. Br J Psychol 2001 Nov;92(Pt 4):659-672 [FREE Full text] [doi: 10.1348/000712601162400] [Medline: 11762867]
28. Kuwahara A, Nishino Y, Ohkubo T, Tsuji I, Hisamichi S, Hosokawa T. Reliability and validity of the Multidimensional Health Locus of Control Scale in Japan: relationship with demographic factors and health-related behavior. Tohoku J Exp Med 2004 May;203(1):37-45 [FREE Full text] [doi: 10.1620/tjem.203.37] [Medline: 15185970]
29. Soravia LM, Schläfli K, Stutz S, Rösner S, Moggi F. Resistance to temptation: the interaction of external and internal control on alcohol use during residential treatment for alcohol use disorder. Alcohol Clin Exp Res 2015 Nov;39(11):2209-2214 [FREE Full text] [doi: 10.1111/acer.12880] [Medline: 26503067]
30. Wallston KA, Stein MJ, Smith CA. Form C of the MHLC scales: a condition-specific measure of locus of control. J Pers Assess 1994 Dec;63(3):534-553. [doi: 10.1207/s15327752jpa6303_10] [Medline: 7844739]
31. Bergvik S, Sørlie T, Wynn R. Coronary patients who returned to work had stronger internal locus of control beliefs than those who did not return to work. Br J Health Psychol 2012 Sep;17(3):596-608 [FREE Full text] [doi: 10.1111/j.2044-8287.2011.02058.x] [Medline: 22151690]
32. Burker EJ, Evon DM, Galanko J, Egan T. Health locus of control predicts survival after lung transplant. J Health Psychol 2005 Sep;10(5):695-704 [FREE Full text] [doi: 10.1177/1359105305055326] [Medline: 16033791]
33. Heidari M, Ghodusi M, Bathaei S, Shakeri K. Self-esteem and Locus of Control in the Initial and Final Stages of Drug Withdrawal Among Addicts Attending Rehabilitation Centers. Addictive Disorders & Their Treatment 2018;17(2):92-97 [FREE Full text] [doi: 10.1097/adt.0000000000000128]
34. Caputo A. Addiction, locus of control and health status: A study on patients with substance use disorder in recovery settings. Journal of Substance Use 2019 Jun 20;24(6):609-613 [FREE Full text] [doi: 10.1080/14659891.2019.1632948]
35. Rowe JL, Montgomery GH, Duberstein PR, Bovbjerg DH. Health locus of control and perceived risk for breast cancer in healthy women. Behav Med 2005;31(1):33-40 [FREE Full text] [doi: 10.3200/BMED.31.1.33-42] [Medline: 16078524]
36. White J, Puckett F, Dutta A, Hayes S, Kundu MM, Johnson E. The Relationship of Multidimensional Health Locus of Control and Attitude toward HIV/AIDS: College Students' Perspectives. Journal of Rehabilitation 2011;77(2):12-18 [FREE Full text]
37. Grotz M, Hapke U, Lampert T, Baumeister H. Health locus of control and health behaviour: results from a nationally representative survey. Psychol Health Med 2011 Mar;16(2):129-140 [FREE Full text] [doi: 10.1080/13548506.2010.521570] [Medline: 21328142]
38. Oaten M, Cheng K. Academic Examination Stress Impairs Self–Control. Journal of Social and Clinical Psychology 2005 Mar;24(2):254-279 [FREE Full text] [doi: 10.1521/jscp.24.2.254.62276]
39. Abay KA, Blalock G, Berhane G. Locus of control and technology adoption in developing country agriculture: Evidence from Ethiopia. Journal of Economic Behavior & Organization 2017 Nov;143:98-115 [FREE Full text] [doi: 10.1016/j.jebo.2017.09.012]
40. Koo D. The moderating role of locus of control on the links between experiential motives and intention to play online games. Computers in Human Behavior 2009 Mar;25(2):466-474 [FREE Full text] [doi: 10.1016/j.chb.2008.10.010]

41. Hsia J. The effects of locus of control on university students' mobile learning adoption. J Comput High Educ 2015 Oct 24;28(1):1-17 [FREE Full text] [doi: 10.1007/s12528-015-9103-8]
42. Hsia J, Chang C, Tseng A. Effects of individuals' locus of control and computer self-efficacy on their e-learning acceptance in high-tech companies. Behaviour & Information Technology 2012 Jul 12;33(1):51-64 [FREE Full text] [doi: 10.1080/0144929x.2012.702284]
43. Hoffman DL, Novak TP, Schlosser AE. Locus of Control, Web use, and Consumer Attitudes toward Internet Regulation. Journal of Public Policy & Marketing 2018 Oct 02;22(1):41-57 [FREE Full text] [doi: 10.1509/jppm.22.1.41.17628]
44. AbuShanab E, Pearson JM, Setterstrom AJ. Internet Banking and Customers' Acceptance in Jordan: The Unified Model's Perspective. CAIS 2010;26:23 [FREE Full text] [doi: 10.17705/1cais.02623]
45. Park N, Kim Y, Shon HY, Shim H. Factors influencing smartphone use and dependency in South Korea. Computers in Human Behavior 2013 Jul;29(4):1763-1770 [FREE Full text] [doi: 10.1016/j.chb.2013.02.008]
46. Drennan J, Kennedy J, Pisarski A. Factors Affecting Student Attitudes Toward Flexible Online Learning in Management Education. The Journal of Educational Research 2005 Jul;98(6):331-338 [FREE Full text] [doi: 10.3200/joer.98.6.331-338]
47. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
48. Bawack RE, Kala Kamdjoug JR. Adequacy of UTAUT in clinician adoption of health information systems in developing countries: The case of Cameroon. Int J Med Inform 2018 Jan;109:15-22. [doi: 10.1016/j.ijmedinf.2017.10.016] [Medline: 29195701]
49. Birch A, Irvine V. Preservice teachers' acceptance of ICT integration in the classroom: applying the UTAUT model. Educational Media International 2009 Dec;46(4):295-315 [FREE Full text] [doi: 10.1080/09523980903387506]
50. Kohnke A, Cole ML, Bush R. Incorporating UTAUT Predictors for Understanding Home Care Patients' and Clinician's Acceptance of Healthcare Telemedicine Equipment. Journal of Technology Management & Innovation 2014 Jul;9(2):29-41 [FREE Full text] [doi: 10.4067/s0718-27242014000200003]
51. Chong AY. Predicting m-commerce adoption determinants: A neural network approach. Expert Systems with Applications 2013 Feb;40(2):523-530 [FREE Full text] [doi: 10.1016/j.eswa.2012.07.068]
52. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
53. Morosan C, DeFranco A. It's about time: Revisiting UTAUT2 to examine consumers' intentions to use NFC mobile payments in hotels. International Journal of Hospitality Management 2016 Feb;53:17-29 [FREE Full text] [doi: 10.1016/j.ijhm.2015.11.003]
54. Zhou T, Lu Y, Wang B. Integrating TTF and UTAUT to explain mobile banking user adoption. Computers in Human Behavior 2010 Jul;26(4):760-767 [FREE Full text] [doi: 10.1016/j.chb.2010.01.013]
55. Fong LHN, Lam LW, Law R. How locus of control shapes intention to reuse mobile apps for making hotel reservations: Evidence from chinese consumers. Tourism Management 2017 Aug;61:331-342 [FREE Full text] [doi: 10.1016/j.tourman.2017.03.002]
56. Venugopal P, Priya S, Manupati V, Varela ML, Machado J, Putnik G. Impact of UTAUT Predictors on the Intention and Usage of Electronic Health Records and Telemedicine from the Perspective of Clinical Staffs. In: Machado J, Soares F, Veiga G, editors. Innovation, Engineering and Entrepreneurship. HELIX 2018. Lecture Notes in Electrical Engineering, vol 505. Cham: Springer; 2019:172-177.
57. Tavares J, Oliveira T. New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
58. Lee JY, Wong CP, Lee SWH. m-Health views and perception among Malaysian: findings from a survey among individuals living in Selangor. Mhealth 2020;6:6 [FREE Full text] [doi: 10.21037/mhealth.2019.09.16] [Medline: 32190617]
59. Samiei V, Wan Puteh SE, Abdul Manaf MR, Abdul Latip K, Ismail A. Are Malaysian Diabetic Patients Ready to Use The New Generation of Health Care Service Delivery? A Telehealth Interest Assessment. Malays J Med Sci 2016 Mar;23(2):44-52 [FREE Full text] [Medline: 27547114]
60. Zailani S, Gilani MS, Nikbin D, Iranmanesh M. Determinants of telemedicine acceptance in selected public hospitals in Malaysia: clinical perspective. J Med Syst 2014 Sep;38(9):111. [doi: 10.1007/s10916-014-0111-4] [Medline: 25038891]
61. Zamani ZA, Nasir R, Yusooff F. Perceptions towards online counseling among counselors in Malaysia. Procedia - Social and Behavioral Sciences 2010;5:585-589 [FREE Full text] [doi: 10.1016/j.sbspro.2010.07.146]
62. Burroughs JE, Glen Mick D. Exploring Antecedents and Consequences of Consumer Creativity in a Problem-Solving Context. J Consum Res 2004 Sep 01;31(2):402-411 [FREE Full text] [doi: 10.1086/422118]
63. Schreier M, Prügl R. Extending Lead-User Theory: Antecedents and Consequences of Consumers' Lead Userness. J Product Innovation Man 2008 Jul;25(4):331-346 [FREE Full text] [doi: 10.1111/j.1540-5885.2008.00305.x]
64. Slade EL, Dwivedi YK, Piercy NC, Williams MD. Modeling Consumers' Adoption Intentions of Remote Mobile Payments in the United Kingdom: Extending UTAUT with Innovativeness, Risk, and Trust. Psychol. Mark 2015 Jul 07;32(8):860-873 [FREE Full text] [doi: 10.1002/mar.20823]

65. Hwang Y. Investigating enterprise systems adoption: uncertainty avoidance, intrinsic motivation, and the technology acceptance model. European Journal of Information Systems 2017 Dec 19;14(2):150-161 [FREE Full text] [doi: 10.1057/palgrave.ejis.3000532]
66. Potosky D, Bobko P. A model for predicting computer experience from attitudes toward computers. Journal of Business and Psychology 2001;15(3):391-404 [FREE Full text]
67. Charlton JP. Measuring perceptual and motivational facets of computer control: the development and validation of the computing control scale. Computers in Human Behavior 2005 Sep;21(5):791-815 [FREE Full text] [doi: 10.1016/j.chb.2004.02.027]
68. Spector PE. Locus of Control and Social Influence Susceptibility: Are Externals Normative or Informational Conformers? The Journal of Psychology 2010 Jul 02;115(2):199-201 [FREE Full text] [doi: 10.1080/00223980.1983.9915436]
69. Avtgis TA. Locus of Control and Persuasion, Social Influence, and Conformity: A Meta-Analytic Review. Psychol Rep 2016 Sep 01;83(3):899-903 [FREE Full text] [doi: 10.2466/pr0.1998.83.3.899]
70. Phares EJ. Locus of control in personality. Morristown, NJ: General Learning Press; 1976.
71. Spector PE. Behavior in organizations as a function of employee's locus of control. Psychological Bulletin 1982 May;91(3):482-497 [FREE Full text] [doi: 10.1037/0033-2909.91.3.482]
72. Burton-Jones A, Hubona GS. The mediation of external variables in the technology acceptance model. Information & Management 2006 Sep;43(6):706-717 [FREE Full text] [doi: 10.1016/j.im.2006.03.007]
73. Igbaria M. The effects of self-efficacy on computer usage. Omega 1995 Dec;23(6):587-605 [FREE Full text] [doi: 10.1016/0305-0483(95)00035-6]
74. Rahi S, Othman Mansour MM, Alghizzawi M, Alnaser FM. Integration of UTAUT model in internet banking adoption context. JRIM 2019 Aug 12;13(3):411-435 [FREE Full text] [doi: 10.1108/jrim-02-2018-0032]
75. Cao Q, Niu X. Integrating context-awareness and UTAUT to explain Alipay user adoption. International Journal of Industrial Ergonomics 2019 Jan;69:9-13 [FREE Full text] [doi: 10.1016/j.ergon.2018.09.004]
76. Chin WW, Marcolin BL, Newsted PR. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. Information Systems Research 2003 Jun;14(2):189-217 [FREE Full text] [doi: 10.1287/isre.14.2.189.16018]
77. Ringle CM, Sarstedt M, Mitchell R, Gudergan SP. Partial least squares structural equation modeling in HRM research. The International Journal of Human Resource Management 2018 Jan 07;31(12):1617-1643 [FREE Full text] [doi: 10.1080/09585192.2017.1416655]
78. Hair J, Risher J, Sarstedt M, Ringle C. When to use and how to report the results of PLS-SEM. EBR 2019 Jan 14;31(1):2-24 [FREE Full text] [doi: 10.1108/ebr-11-2018-0203]
79. Hair JF, Hult GTM, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Thousand Oaks, CA: Sage Publications Inc; 2016.
80. Chin WW. The partial least squares approach to structural equation modeling. In: Marcoulides GA, editor. Modern Methods for Business Research. Mahwah, NJ: Lawrence Erlbaum Associates; 1998:295-336.
81. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J. of the Acad. Mark. Sci 2014 Aug 22;43(1):115-135 [FREE Full text] [doi: 10.1007/s11747-014-0403-8]
82. Taffesse A, Tadesse F. Pathways Less Explored—Locus of Control and Technology Adoption. Journal of African Economies 2017;26(suppl_1):i36-i72 [FREE Full text] [doi: 10.1093/jae/ejx013]
83. Helmer SM, Krämer A, Mikolajczyk RT. Health-related locus of control and health behaviour among university students in North Rhine Westphalia, Germany. BMC Res Notes 2012 Dec 29;5:703 [FREE Full text] [doi: 10.1186/1756-0500-5-703] [Medline: 23273039]
84. Calnan M. Control over health and patterns of health-related behaviour. Social Science & Medicine 1989 Jan;29(2):131-136 [FREE Full text] [doi: 10.1016/0277-9536(89)90161-5] [Medline: 2749298]
85. Mirzania M, Khajavi A, Kharazmi A, Moshki M. Health literacy and quality of life among Iranian pregnant women: The mediating role of health locus of control. Med J Islam Repub Iran 2020 Oct 30;34:161 [FREE Full text] [doi: 10.47176/mjiri.34.161] [Medline: 33816360]
86. Tsai T, Lee SD, Tsai Y. Explaining selected health behaviors in a national sample of Taiwanese adults. Health Promot Int 2015 Sep 15;30(3):563-572 [FREE Full text] [doi: 10.1093/heapro/dat085] [Medline: 24342796]
87. Cohen M, Azaiza F. Health-promoting behaviors and health locus of control from a multicultural perspective. Ethn Dis 2007;17(4):636-642. [Medline: 18072372]
88. Poortinga W, Dunstan FD, Fone DL. Health locus of control beliefs and socio-economic differences in self-rated health. Prev Med 2008 Apr;46(4):374-380 [FREE Full text] [doi: 10.1016/j.ypmed.2007.11.015] [Medline: 18177930]
89. Zhou LL, Owusu-Marfo J, Asante Antwi H, Antwi MO, Kachie ADT, Ampon-Wireko S. "Assessment of the social influence and facilitating conditions that support nurses' adoption of hospital electronic information management systems (HEIMS) in Ghana using the unified theory of acceptance and use of technology (UTAUT) model". BMC Med Inform Decis Mak 2019 Nov 21;19(1):230 [FREE Full text] [doi: 10.1186/s12911-019-0956-z] [Medline: 31752840]

90. Bozan K, Davey B, Parker K. Social Influence on Health IT Adoption Patterns of the Elderly: An Institutional Theory Based Use Behavior Approach. Procedia Computer Science 2015;63:517-523 [FREE Full text] [doi: 10.1016/j.procs.2015.08.378]
91. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 09;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]

Abbreviations

AVE: average variance extracted
HLOC: health locus of control
HTMT: heterotrait-monotrait
ICT: information and communications technology
LOC: locus of control
mHealth: mobile health
PLS-SEM: partial least squares structural equation modeling
UTAUT: Unified Theory of Acceptance and Use of Technology
VIF: variance inflation factor

Edited by R Kukafka; submitted 19.02.21; peer-reviewed by KC Wong, K Rawlings; comments to author 25.05.21; revised version received 22.06.21; accepted 05.07.21; published 29.12.21.

Please cite as:
Ahadzadeh AS, Wu SL, Ong FS, Deng R
The Mediating Influence of the Unified Theory of Acceptance and Use of Technology on the Relationship Between Internal Health Locus of Control and Mobile Health Adoption: Cross-sectional Study
J Med Internet Res 2021;23(12):e28086
URL: https://www.jmir.org/2021/12/e28086
doi:10.2196/28086
PMID:34964718

Original Paper

Assessing the Implementation and Effectiveness of the Electronic Patient-Reported Outcome Tool for Older Adults With Complex Care Needs: Mixed Methods Study

Carolyn Steele Gray[1,2], MSc, PhD; Edward Chau[2], MSc; Farah Tahsin[2], MA; Sarah Harvey[3], PhD; Mayura Loganathan[4,5], MD; Brian McKinstry[6], MD, PhD; Stewart W Mercer[6], MBChB, PhD; Jason Xin Nie[7], MSc; Ted E Palen[8], MD, PhD; Tim Ramsay[9,10], PhD; Kednapa Thavorn[9,10], PhD; Ross Upshur[1,5], MD, PhD; Walter P Wodchis[2,7], PhD

[1]Bridgepoint Collaboratory for Research and Innovation, Lunenfeld-Tanenebaum Research Institute, Sinai Health, Toronto, ON, Canada
[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Logibec Inc (QoC Health Inc), Toronto, ON, Canada
[4]Ray D Wolfe Department of Family Medicine, Mount Sinai Hospital, Toronto, ON, Canada
[5]Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[6]Usher Institute, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom
[7]Institute for Better Health, Trillium Health Partners, Mississauga, ON, Canada
[8]Institute for Health Research, Kaiser Permanente Colorado, Denver, CO, United States
[9]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[10]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Carolyn Steele Gray, MSc, PhD
Bridgepoint Collaboratory for Research and Innovation
Lunenfeld-Tanenebaum Research Institute
Sinai Health
1 Bridgepoint Drive
Toronto, ON, M4M 2B5
Canada
Phone: 1 4168047100
Email: Carolyn.SteeleGray@sinaihealth.ca

Abstract

Background: Goal-oriented care is being adopted to deliver person-centered primary care to older adults with multimorbidity and complex care needs. Although this model holds promise, its implementation remains a challenge. Digital health solutions may enable processes to improve adoption; however, they require evaluation to determine feasibility and impact.

Objective: This study aims to evaluate the implementation and effectiveness of the electronic Patient-Reported Outcome (ePRO) mobile app and portal system, designed to enable goal-oriented care delivery in interprofessional primary care practices. The research questions driving this study are as follows: Does ePRO improve quality of life and self-management in older adults with complex needs? What mechanisms are likely driving observed outcomes?

Methods: A multimethod, pragmatic randomized controlled trial using a stepped-wedge design and ethnographic case studies was conducted over a 15-month period in 6 comprehensive primary care practices across Ontario with a target enrollment of 176 patients. The 6 practices were randomized into either early (3-month control period; 12-month intervention) or late (6-month control period; 9-month intervention) groups. The primary outcome measure of interest was the Assessment of Quality of Life-4D (AQoL-4D). Data were collected at baseline and at 3 monthly intervals for the duration of the trial. Ethnographic data included observations and interviews with patients and providers at the midpoint and end of the intervention. Outcome data were analyzed using linear models conducted at the individual level, accounting for cluster effects at the practice level, and ethnographic data were analyzed using qualitative description and framework analysis methods.

Results: Recruitment challenges resulted in fewer sites and participants than expected; of the 176 target, only 142 (80.6%) patients were identified as eligible to participate because of lower-than-expected provider participation and fewer-than-expected

patients willing to participate or perceived as ready to engage in goal-setting. Of the 142 patients approached, 45 (32%) participated. Patients set a variety of goals related to self-management, mental health, social health, and overall well-being. Owing to underpowering, the impact of ePRO on quality of life could not be definitively assessed; however, the intervention group, ePRO plus usual care (mean 15.28, SD 18.60) demonstrated a nonsignificant decrease in quality of life (t_{24}=−1.20; *P*=.24) when compared with usual care only (mean 21.76, SD 2.17). The ethnographic data reveal a complex implementation process in which the meaningfulness (or coherence) of the technology to individuals' lives and work acted as a key driver of adoption and tool appraisal.

Conclusions: This trial experienced many unexpected and significant implementation challenges related to recruitment and engagement. Future studies could be improved through better alignment of the research methods and intervention to the complex and diverse clinical settings, dynamic goal-oriented care process, and readiness of provider and patient participants.

Trial Registration: ClinicalTrials.gov NCT02917954; https://clinicaltrials.gov/ct2/show/NCT02917954

(J Med Internet Res 2021;23(12):e29071) doi:10.2196/29071

KEYWORDS

older adults; goal-oriented care; quality of life; self-management; primary care; eHealth; pragmatic trial; mobile phone

Introduction

Background

The rising population of older adults with multimorbidity and complex care needs requires that health systems adjust to meet this new demand [1,2]. Complex care needs of patients go beyond multimorbidity alone, as these individuals will experience biopsychosocial challenges and barriers that make it more difficult for them to manage their multiple chronic physical and mental illnesses [3,4]. Increasingly, digital health solutions are being adopted to support this patient population through tools that enable medication management [5], information sharing [6], care planning [7], chronic disease management and monitoring [8,9], and virtual care tools, particularly since the onset of the COVID-19 pandemic [10]. Of particular use to older adults with complex care needs are solutions that enable person-centered and holistic care delivery to better address their multiple health and social care needs [3,11-17]. Although a person-centered approach has been identified as a priority [16], organizations and providers continue to struggle with how to put the vision of person-centered care into practice [18].

Person-centered care may be operationalized by adopting a goal-oriented care (GOC) approach that involves eliciting patient-identified goals to drive care planning and decision-making [13,14,19,20]. Effectively, this model of care shifts from asking a patient "What is the matter with you?" to "What matters to you?" [21] From a patient perspective, GOC represents a more meaningful and holistic approach to care and decision-making [22]. Emerging studies of GOC report reduced treatment burden for patients with multiple chronic conditions [23] and reductions in acute inpatient days and mortality [24]. The pragmatic trial of the Health TAPESTRY program, a digitally enabled, community-based GOC program, evaluated the program's impact on goal attainment, self-efficacy, quality of life, optimal aging, social support, empowerment, physical activity, falls, and access. The Health TAPESTRY trial demonstrated a shift from reactive to proactive care [25]; however, similar to many other studies of person-centered care [26,27], Health TAPESTRY did not demonstrate an impact on patient outcomes.

Among the challenges in evaluating an approach such as GOC, in particular a digitally enabled GOC model, is that it is a complex intervention that is delivered to a complex patient population within a complex system. Conventional methods, such as randomized controlled trials, that apply rigid methods and rely on assumptions of linear cause-effect perspectives [28] may result in *controlling* for the variables that we need to capture [29]. Greenhalgh and Papoutsi [28] instead suggest methods that adopt a systems mindset that allows for adaptability, iteration, and design-thinking better suited to capturing "changing interrelationships between parts of the system." The evaluation presented in this paper adopts this systems mindset to evaluate the electronic Patient-Reported Outcome (ePRO) tool, a novel mobile device and a linked portal system that enables GOC delivery to older adults with complex care needs receiving care in interdisciplinary primary care practices. This evaluation is the latest iteration of a multiphase developmental evaluation of ePRO that took place in Ontario, Canada, from April 2018 to June 2019.

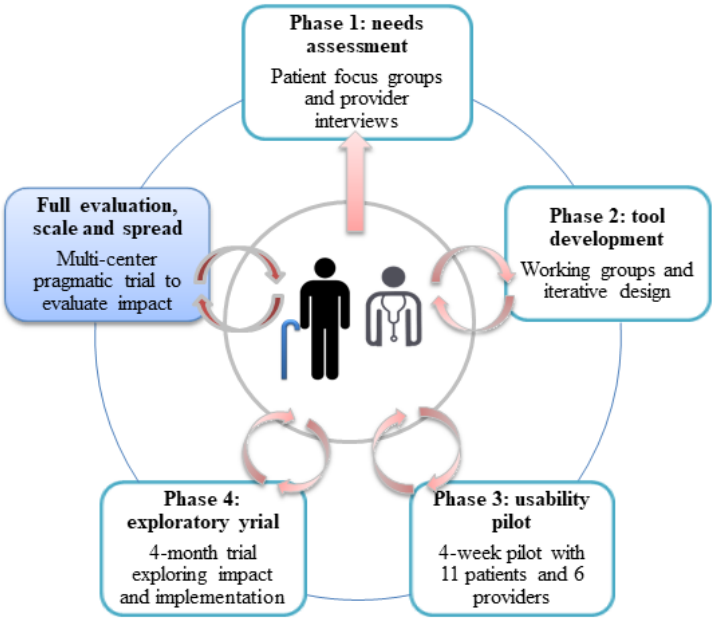
Objective and Research Questions

This developmental evaluation incorporates a pragmatic, stepped-wedge, cluster trial with embedded ethnographic case studies, building on previous stages of design, development, and testing [30-33] (see Figure 1 for a visual representation of how this work builds on previous stages). This study expands the findings from our exploratory trial [33] as a means to engage in what Tsoukas terms "conjunctive theorizing to generate rich pictures of complex phenomena by drawing together different kinds of data from multiple sources" [34]. This work was guided by the following research questions:

1. Does ePRO improve quality of life and self-management in older adults with complex needs?
2. What mechanisms are likely driving observed outcomes?

Regarding the first research question, it is hypothesized that the ePRO tool will have a positive impact on quality of life and patients' ability to self-manage.

Figure 1. electronic Patient-Reported Outcome (ePRO) co-design steps.

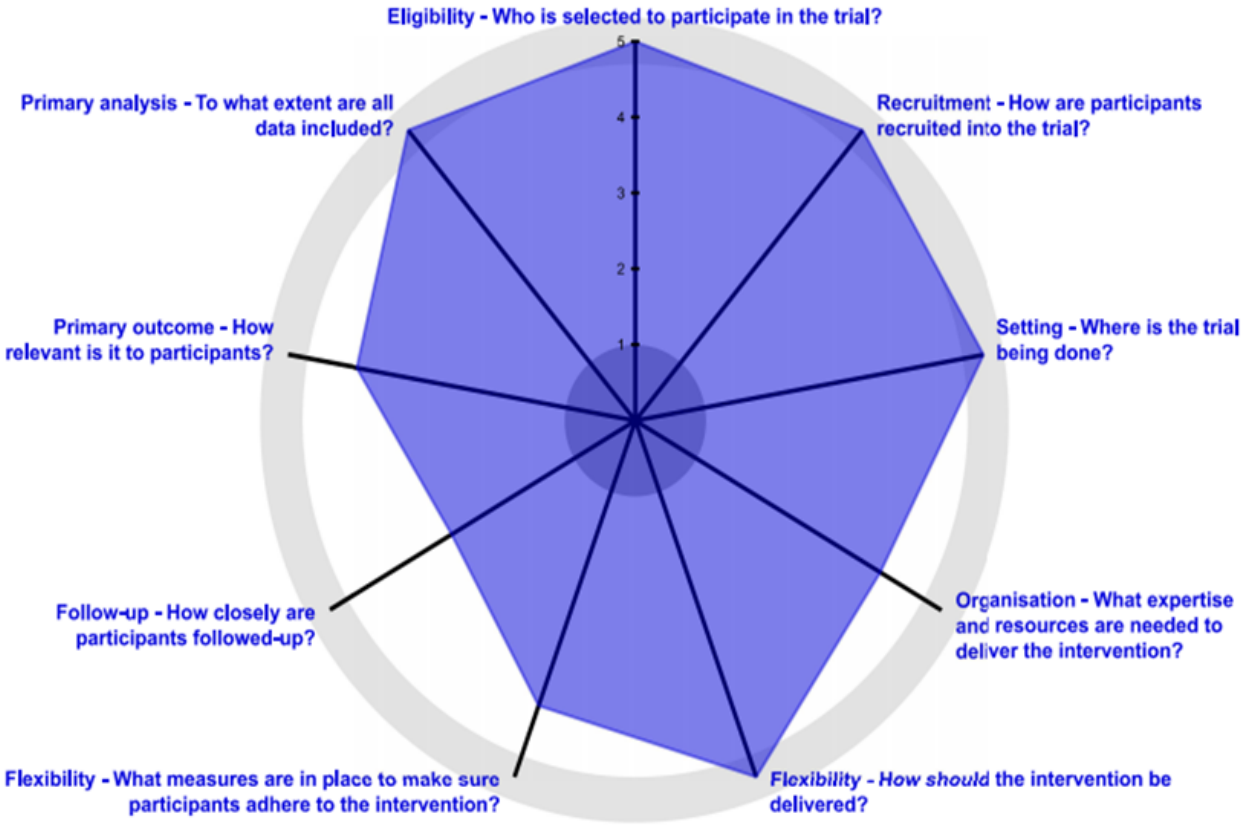


Methods

Design

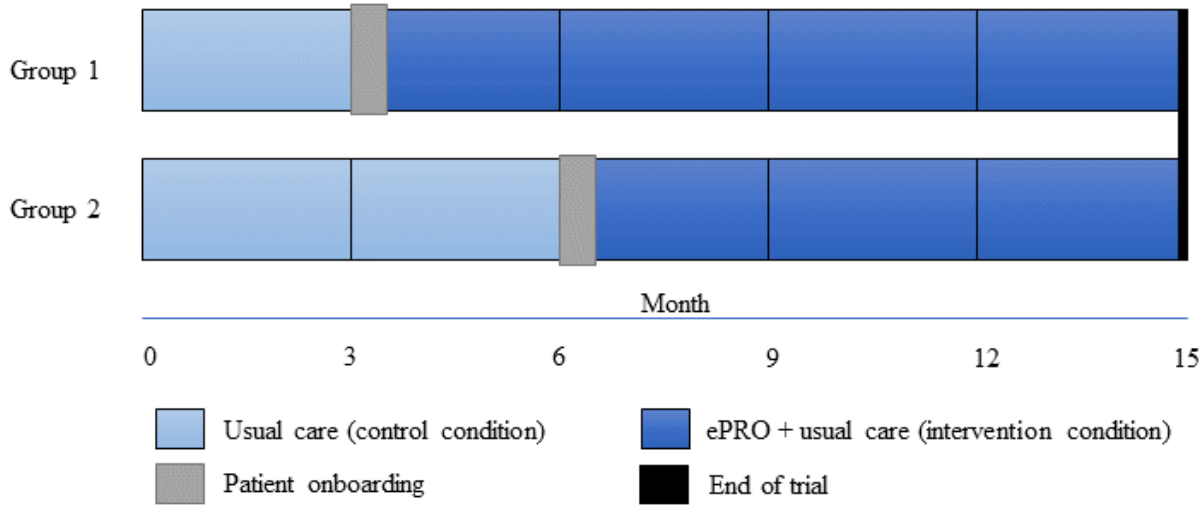
Aligned with Medical Research Council guidelines for evaluating complex interventions [35], a developmental evaluation approach was applied to collect outcome, process, and context measures to support decision-making and technology modifications [36]. A pragmatic, stepped-wedge, cluster randomized trial design was used to assess the effectiveness of the ePRO tool [37]. The Pragmatic Explanatory Continuum Indicator Summary (PRECIS-2) wheel in Figure 2 (see Multimedia Appendix 1 for description of the wheel domains as related to this trial) describes the degree to which the trial represents a pragmatic design. This design was the most feasible and appropriate approach given the nature of the intervention, time, and resources available [38] and the desire to complete in a real-world setting [39]. An embedded ethnographic case study was included, aligned with the methods outlined by Greenhalgh and Swinglehurt [40] for evaluating complex technological innovations. The case studies offer rich contextual and process information that accounts for complex interrelationships between variables that are missed by looking at outcome data alone.

Figure 2. PRECIS-2 (Pragmatic Explanatory Continuum Indicator Summary) Wheel for electronic Patient-Reported Outcome (ePRO) trial.



The trial was conducted in 6 comprehensive primary care practices called Family Health Teams (FHTs) across Ontario, Canada, over a 15-month period. Following the stepped-wedge design, all FHT sites started in the control period where all recruited patients received usual care (no change to their care delivery from the primary care team). A random number generator assigned sites to either the early intervention (n=3 sites) or late intervention (n=3 sites) groups. The early intervention group (group 1) was assigned to the intervention for 12 months after the initial 3-month control period. The FHTs in group 2 were switched to the intervention group for 9 months after a 6-month control period. Figure 3 shows a diagram of the stepped-wedge design.

Figure 3. Stepped-wedge design for electronic Patient-Reported Outcome (ePRO) evaluation.



Intervention: The ePRO Tool

The ePRO tool was developed via multiphase user-centered co-design methods and represents an important divergence from many available systems that are focused on a single disease or are built to enhance existing provider-led models of care. The tool is designed to encourage a shift in the care process toward a person-centered model by enabling the full GOC process, including goal elicitation, ongoing monitoring, and goal modification [41]. Consistent with co-design methods, the tool was iteratively developed with input from patients with complex care needs, caregivers, and a multidisciplinary primary care team [30,31]. The tool has undergone usability testing [32] and an exploratory trial [33]. Findings from these studies were used to update and adapt the tool to user needs and different primary care settings. At the time of the trial, the ePRO tool did not connect to other existing technology systems, such as electronic medical records (EMRs) or other available platforms; however, the system was built so interoperability would be possible (see Multimedia Appendix 2 for wireframes).

Population and Setting

A 2-stage sampling strategy was implemented, first recruiting FHTs, followed by complex patients within each FHT. FHTs in Ontario are similar to Patient-Centered Medical Homes in the United States in that they both seek to provide comprehensive primary care services through a physician-led multidisciplinary team [42]. Working in collaboration with the project's decision-making partner, the Association of Family Health Teams of Ontario (AFHTO; representing all 184 FHTs in Ontario), a multipronged FHT recruitment strategy was pursued, including (1) email invitations sent to AFHTO member sites, (2) a webinar session with AFHTO quality improvement specialists who could identify eligible sites, and (3) an information booth at the annual AFHTO conference (October 2016 in Toronto, Ontario) where study information was shared with delegates. From these avenues, 29 sites expressed interest to be assessed for eligibility, with 6 FHTs agreeing to participate (see the Figure 4 CONSORT flow diagram of site recruitment). As FHTs are geographically diverse, there is no chance of cross-contamination of providers across different sites. The characteristics of the participating sites, as compared with FHTs across Ontario, are summarized in Table 1, and the population densities of the regions are depicted in Figure 5. As can be seen in Figure 5, sites A and F were in rural settings, sites D and E were in urban settings, and sites B and C were medium urban as consistent with Statistics Canada definitions of rurality [43]. Approximately 36% (59/165) of FHTs are located in rural settings.

Providers eligible to participate in the study had to be provided care to patients rostered at the FHT. Providers can be employed full-time, part-time, or casual.

Figure 4. CONSORT (Consolidated Standard of Reporting Trials) flow diagram–Family Health Team (FHT) recruitment.

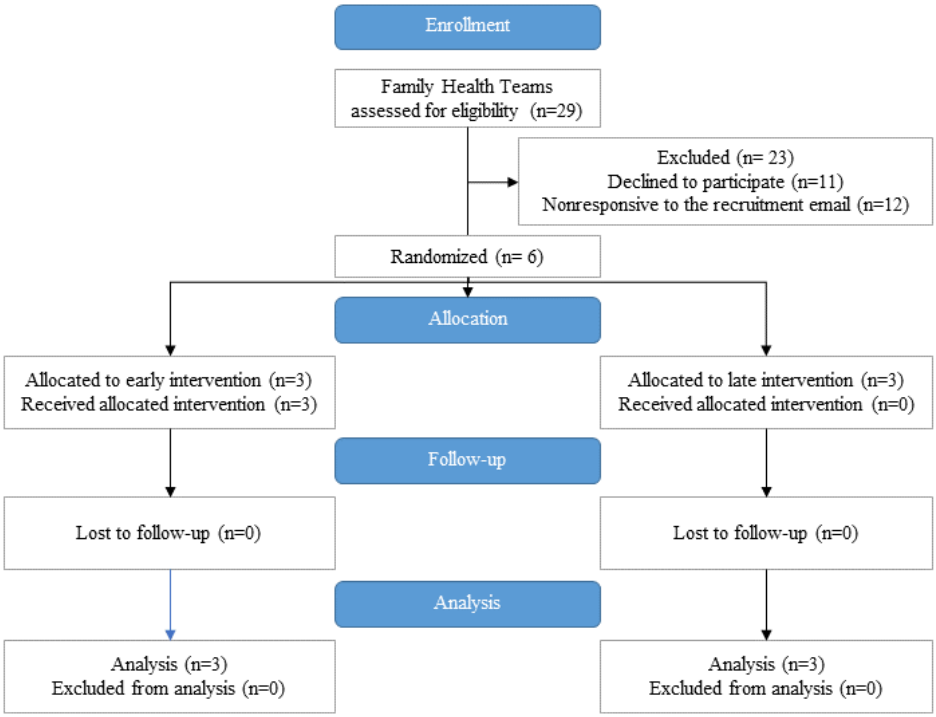


Figure 5. Population density across the regions (persons per square km). Source: Statistics Canada 2016 Census [43].

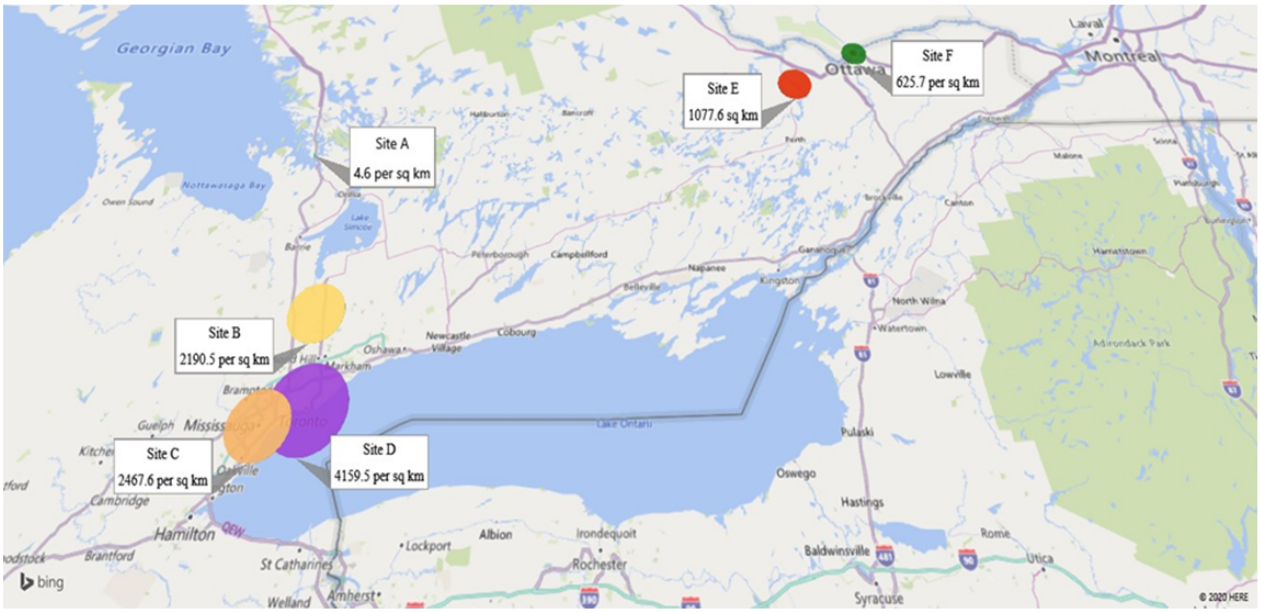


Table 1. Family Health Team characteristics[a].

Characteristics	Group 1						Group 2						Ontario FHTs[b], mean (SD)
	Site A		Site D		Site E		Site B		Site C		Site F		
	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	
Number of providers enrolled in the ePRO[c] study													
Total number of providers	9 (69)	13	4 (22)	18	6 (27)	22	1 (6)	17	2 (17)	12	7 (54)	13	N/A[d]
GPs[e]	0 (0)	4	1 (18)	12	1 (17)	14	1 (19)	11	2 (22)	9	6 (75)	8	13.53 (17.72)[f]
NPs[g]	8 (100)	8	1 (33)	3	4 (67)	6	0 (0)	4	0 (0)	2	1 (0)	—[h]	2.65 (3.04)
RD[i]	1 (100)	1	2 (100)	2	1 (100)	1	0 (0)	1	0 (0)	1	0 (0)	0	1.19 (1.63)
Pharmacist	0 (0)	0	0 (0)	1	0 (0)	1	0 (0)	1	0 (0)	1	0 (0)	0	0.63 (0.99)

[a]Site names were assigned based on the timing of recruitment. Ontario FHT data available from 165 FHT sites.

[b]FHT: Family Health Team.

[c]ePRO: electronic Patient-Reported Outcome.

[d]N/A: not applicable.

[e]eGPs: general practitioners.

[f]Information available from 165 FHTs across Ontario.

[g]NP: nurse practitioner.

[h]Not available.

[i]RD: registered dietitian.

Patient Recruitment and Eligibility Criteria

Patient recruitment followed exploratory trial procedures, using practice EMRs to identify patients aged ≥60 years with ≥10 visits to the FHT within the previous 12 months. This number of visits has been identified as an indicator of complexity in previous studies [44] and guided the recruitment strategy for the exploratory trial [33]. Age 60 years was chosen as a cut-off over 65 years as the study's site leads and primary care knowledge user partners identified that many individuals, particularly those living in rural settings, experience complex care needs at an earlier age. EMR-generated patient lists were given to providers to assess whether these individuals met the additional eligibility criteria: (1) perceived willingness to engage in GOC conversations, (2) ability to use a smartphone or tablet in English or have a caregiver who could do this on their behalf,

(3) capable of providing consent to participate, (4) willing to complete surveys every 3 months thereafter until the trial concluded. Previous studies have identified that the provider knowledge of the patient is often necessary to identify complexity given the high degree of patient variability [38].

Posters describing the study were hung in waiting rooms at sites, and the study was presented at chronic disease management programs that targeted patients with chronic disease and complex care needs for patient self-identification. Some patients self-identified as eligible after presentations at the programs, but none came to the study via posters. Recruitment materials and processes were built on what was learned from the exploratory trial and were reviewed and modified by the project's patient partner. Recruitment occurred during a scheduled office visit or by phone by a research coordinator assigned to that site. Patient and provider consent was obtained before randomization.

The minimum sample size required for the recruitment of sites and patients was determined using closed-form analytic formulas with a power of 80% based on a minimal clinically important difference of our core measure of quality of life (the AQoL-4D) of 0.06 [45], an expected SD in assessment of quality of life (AQoL) of 0.22 [46], an expected intraclass correlation coefficient (ICC) of 0.01 (calculated based on total primary care use over a 1-year period among a 10% sample of the Ontario population, which served here as a proxy measure for patient outcomes), and an expected attrition rate of 10% (rated based on previous studies in similar population groups using similar technology [47]). A minimum sample size of 176 patients was calculated, with a target of recruiting 29 patients per site.

Technology Training

Providers and patients were trained on the tool before switching from the control to the intervention during an onboarding session. Training for providers was done at the clinic level, where groups of providers were presented the technology by a research team member who walked through the technology by setting up goals for a mock patient. Patients were trained one-on-one with a research team member on how to use the mobile device and platform just before their onboarding visit with the provider. Providers and patients were also provided with user manuals [48] and contact information for the research team for technology support.

Data Collection

Context, process, and outcome data were collected via patient-reported surveys, interviews, ethnographic observations, and chart audits. Survey and chart audit data were collected across all 6 sites, whereas qualitative data were collected at the 3 case sites (sites A, E, and F). In total, 4 of 6 agreed to participate as case sites, and 1 dropped out as a case site because of low patient recruitment.

Patient-Reported Surveys

The primary outcome for this study was health-related quality of life measured using the Assessment of Quality of Life–4 Dimensions (AQoL-4D) [49]. The AQoL-4D is a 12-item questionnaire that addresses the activities of daily living, mental health, relationships with others, and physical aspects of a patient's quality of life. AQoL-4D responses were aggregated to generate a raw, continuous score with higher scores indicating a greater quality of life. The secondary outcome, self-management, was measured using the 13-item Patient Activation Measure (PAM), which measures the extent to which a patient is activated in their own care [50]. PAM is considered a valid and reliable metric of patient self-management for older adults with multimorbidity [50-52]. PAM generates a score from 0 to 100, with higher scores indicating greater activation (patients are better able to manage their care) [53]. Outcome data were collected at baseline and every subsequent 3 months until the end of the trial. For 3 patients in group 2 who were enrolled in the study late, outcome data were collected between October 2018 and June 2019 (Multimedia Appendix 3).

Patient and provider demographic information were collected at baseline. A chart review was conducted posttrial to collect missing data in the patient demographic forms, particularly the number of types of chronic conditions and medications.

Interviews

Semistructured interviews were conducted with patients, providers, and managers at case sites at the midpoint (6 months for sites A and E and 4.5 months for site F) and end of the trial. Interviews were conducted by research team members trained in qualitative data collection, with initial interviews conducted in pairs to ensure consistency in the approach. The interview guide was developed to capture the experiences of patients, providers, and managers using the tool or engaged with the trial. Probes were used to delve into implementation factors found to be important to the intervention in the exploratory trial, for example, patient-provider relationships and team environment [33] (Multimedia Appendix 4). Interviews were conducted in person (with one follow-up midpoint interview conducted over the phone), lasted between 20 and 60 minutes, and were audio-recorded and transcribed verbatim.

Ethnographic Observation

Ethnographic observations of case sites occurred at multiple points throughout the study, mainly when conducting other activities such as training, patient onboarding, and interviews. At these points, a member of the research team would observe the clinic visits between the patient and provider. Providers were also encouraged to inform the team when patients were coming in for visits so that additional *ad hoc* observations could be conducted; however, no such invitations occurred. Field memos were taken during and immediately after the observation periods. Field note guides helped research staff attend to contexts and processes anticipated to be relevant based on findings from the exploratory trial. Observations were conducted by research coordinators who had graduate training in qualitative health services methods or were provided training by the project lead in the approach. For coordinators, newer to the method observation debriefs and field memo reviews were conducted by the lead to provide ongoing training and skill building.

Use Logs

Use logs from the ePRO tool were used to track tool use and the types of goals set by the participants. Goals were categorized

into types using qualitative content analysis. Tool use was determined by the number of interactions defined as any log-in or data entered into the system; participants completing one interaction in a given month were considered *active* that month. The total number of active participants was calculated monthly to categorize participants into long-term (using the app for 3 or more months), short-term (using the app for <3 months), or nonuser (participants who did not use the app after initial onboarding) groups. The 3-month cut-off was consistent with previous mobile health (mHealth) clinical trials [54]. Use categories helped to interpret the qualitative and quantitative findings.

Data Analyses

Statistical Analysis

Descriptive statistics were calculated for the cohort stratified by groups of FHTs using counts and mean (SD) values for categorical and continuous variables, respectively. To estimate the degree to which the ePRO tool plus usual care affects health-related quality of life and self-management relative to usual care alone, linear models were fitted with exposure identified by a fixed-effect ordinal variable of calendar time (accounting for staggered implementation) and adjusting for clustering at the FHT site level [55]. The primary effect estimates are summarized as the mean differences for continuous outcomes. Each comparison was evaluated using a 2-sided test at a nominal significance level of α=.05. Statistical analyses of outcome data were performed using the intention-to-treat principle. All descriptive analyses and multilevel modeling were completed using Stata 15.1 statistical software (StataCorp LLC). Owing to the size of the data set, mixed effects models that included covariates such as age, sex, income level, rurality, chronic disease management, and number of chronic conditions could not be included in the modeling.

While missingness in cluster randomized trials in primary care can be handled via multiple imputation methods, using any such imputation to estimate the absence of data points in the cohort was deemed inappropriate owing to the high degree of missingness [56,57]. Aligned with the intention-to-treat approach, individuals were not excluded from the analysis based on their nonresponses to the survey; only variables were excluded, not individuals.

Interview and Observation Data

Interview and observational data were used to address the second research question and were analyzed using inductive qualitative descriptive [58] and narrative descriptive methods [59], with separate analyses conducted for patient and provider interviews. Manager interviews were coded with provider data as they were asked similar questions and addressed many of the same implementation constructs. Consistent with this method, codes that described the dominant themes within participant groups were identified. The coding was conducted by researcher pairs trained using qualitative methods to support the validation. Observational memos were coded with patient interviews and were also reviewed as part of the analytic process to provide context information where appropriate to guide interpretation. All team members involved in qualitative data collection and analysis were trained to attend to reflexivity in their approach and all kept fulsome analytic memos to track their own positionality with regard to the qualitative analysis.

To support directed analysis for the purposes of this evaluation, a deductive approach was used to map descriptive codes to Normalization Process Theory (NPT) [60] to understand implementation mechanisms. Exploratory trial findings suggest that NPT is a likely theory of change that underpins this intervention [33]. NPT suggests that new processes become embedded as part of actors' routines through the social production of work, enabled through 4 generative mechanisms: coherence, cognitive participation, collective action, and reflexive monitoring (Multimedia Appendix 5 offers descriptions of the concepts). These 4 NPT constructs were applied to descriptively coded patient, provider, and manager interviews and observational data, and cross-referenced with patient user groups (long, short, and non) and case site characteristics to generate insights regarding factors that drove implementation and outcomes. Data coded to relevant themes were extracted and organized using tables using a framework analysis approach [61] to identify patterns and trends. The research team reviewed the tables as part of the qualitative validation (supporting credibility and trustworthiness). NVivo 11 software (QSR International, version 11, 2015) was used to manage data in the initial coding phase, and Microsoft Excel and Word files were used to organize data for the framework analysis.

Integrating Quantitative and Qualitative Data

Integration of quantitative and qualitative data sets followed a *convergent design* that involves collecting all sources of data concurrently, separately analyzing data, and then comparing results through interpretation and discussion of findings [62,63].

Ethics

Research ethics approval was granted by the University of Toronto's Health Sciences Research Ethics Board (#33944) and the ethics committees of all participating practices.

Results

Participant Recruitment

Although the study design target was 176 patients, only 142 (80.6%) were identified as eligible and approached. Of the 142 approached, 46 (32.4%) consented to participate. One participant withdrew before any data collection, leaving 45 (31.6%) participants. This relatively low acceptance rate is an additional challenge. Patient-reported reasons for not participating included perceiving that they did not have complex or chronic conditions, lack of time, perceived conflicts with other life responsibilities (eg, planned vacations and travel), feeling as though they did not have a goal to work on, or were uninterested in this research. In total, 7% (3/45) patients dropped out of the study (1) because of a decline in health condition, making it difficult to participate, and (2) because of loss of interest in participating. Figure 6 shows the CONSORT flow diagram depicting patient recruitment. Figure 7 offers a summary of the number of patient and provider participants per site.

Figure 6. CONSORT (Consolidated Standard of Reporting Trials) flow diagram of patient recruitment arranged by group.

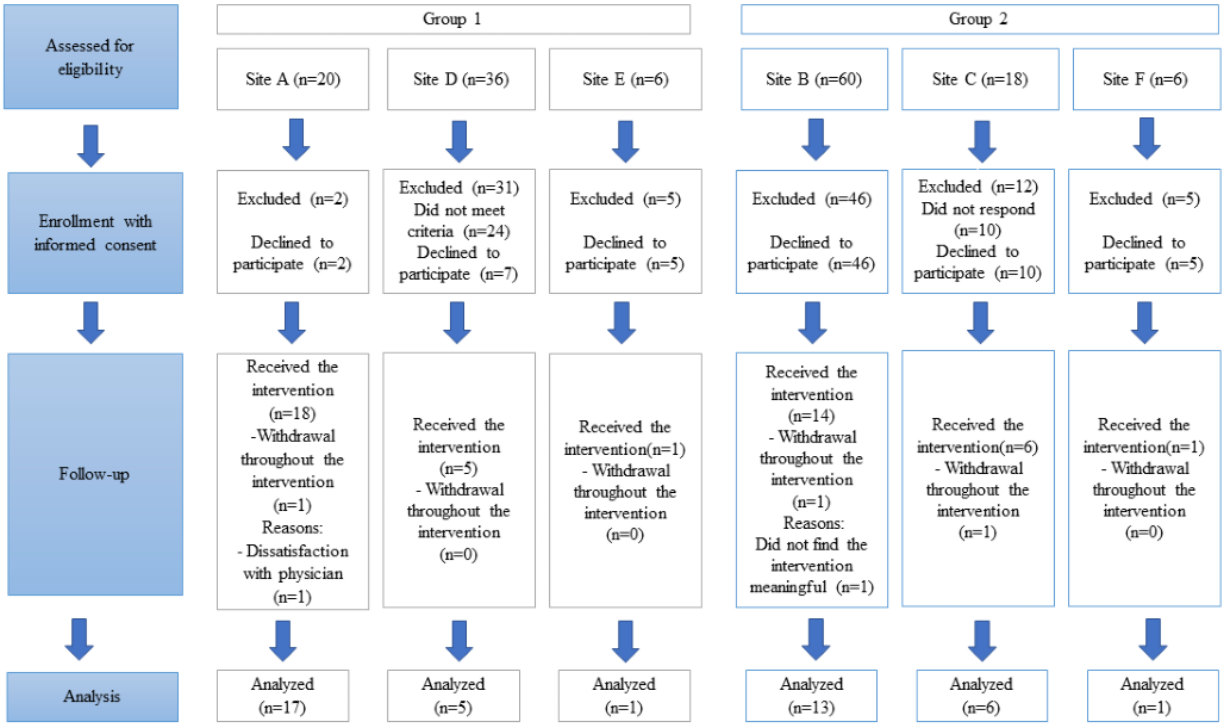
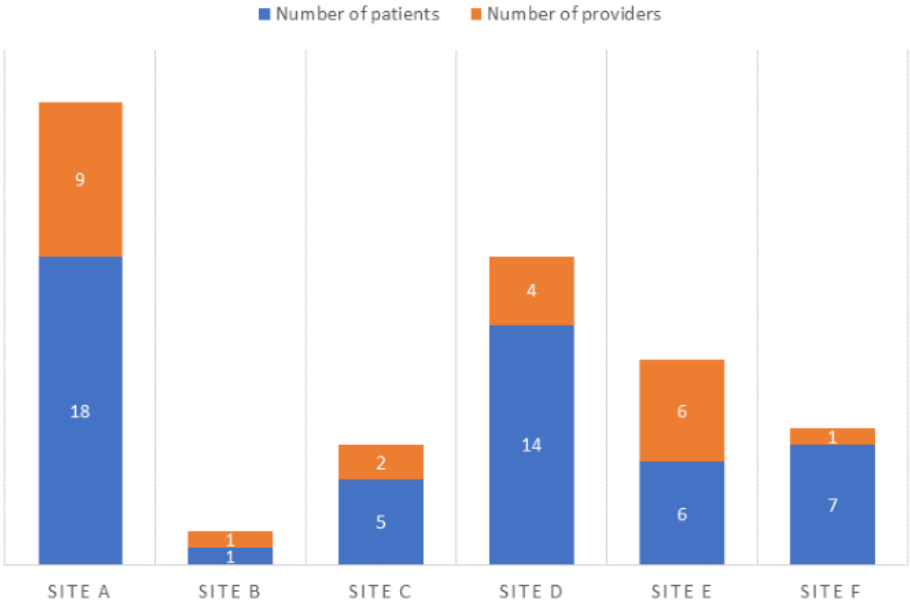


Figure 7. Number of providers and patients participating at each site.



Participant Characteristics and Goals Set

Patient demographics are summarized in Table 2 and presented by randomized groups to allow for between-group comparisons. There was a statistically significant difference in the rurality and socioeconomic status between the groups.

Patients set a variety of goals related to self-management, mental health, social health, and overall well-being and self-care (Multimedia Appendix 6). Patient-provider pairs varied in terms of the degree of specificity of the goals they set, ranging from highly specific goals (eg, walking 20 minutes 3 times per week or losing 10 pounds) to more general goals (eg, reducing meat consumption or getting more sleep).

Table 2. Baseline characteristics of the cohort of Family Health Team patients with complex chronic diseases and disabilities (n=44)[a].

Characteristics	Group 1 (site A, site D, and site E; n=23)	Group 2 (site B, site C, and site F; n=21)	*P* value
Age (years), mean (SD)	68.65 (7.10)	71.98 (6.20)	.08
Sex, n (%)			.07
Female	15 (65.22)	7 (33.33)	
Male	8 (34.78)	14 (66.67)	
Place of residence, n (%)			.08
Urban	9 (39.13)	14 (66.67)	
Rural	14 (60.87)	7 (33.33)	
Living alone, n (%)			.36
Yes	10 (43.48)	6 (28.57)	
No	13 (56.52)	15 (71.43)	
Born in Canada, n (%)			.17
Yes	19 (82.62)	13 (61.90)	
No	4 (17.39)	8 (38.10)	
Family income [CAD $ (US $)][b], n (%)			.04
0-29,000 (0-24,199)	7 (30.43)	1 (4.76)	
30,000-59,000 (24,200-48,398)	7 (30.43)	5 (23.81)	
60,000-89,000 (48,399-72,598)	2 (8.70)	8 (38.10)	
>90,000 (>72,599)	7 (30.43)	7 (33.33)	
Education[c], n (%)			.02
Less than high school	4 (17.39)	1 (4.76)	
High school	4 (17.39)	1 (4.76)	
Some college or university	9 (39.13)	4 (19.05)	
University (undergraduate or graduate)	6 (26.09)	15 (71.43)	
Ethnicity, n (%)			.43
East Asian	0 (0.00)	1 (4.76)	
South Asian	1 (4.35)	0 (0.00)	
Metis	0 (0.00)	1 (4.76)	
White (North American or European)	21 (91.30)	17 (80.95)	
Mixed heritage	1 (4.35)	2 (9.52)	
Chronic disease management program, n (%)			>.99
Yes	6 (26.09)	2 (9.52)	
No	1 (4.40)	1 (5.00)	
Missing	16 (69.57)	18 (85.71)	
Total number of chronic conditions, mean (SD)	4.21 (2.00)	3.20 (2.00)	<.001
Chronic conditions diagnoses, n (%)			
Arthritis	7 (30.43)	2 (9.52)	—[d]
Asthma	5 (21.74)	3 (14.30)	—
Atrial fibrillation	1 (4.40)	2 (9.52)	—
Cancer	8 (35.00)	3 (14.30)	—
Chronic obstructive pulmonary disease	10 (44.00)	2 (9.52)	—

Characteristics	Group 1 (site A, site D, and site E; n=23)	Group 2 (site B, site C, and site F; n=21)	*P* value
Congestive heart failure	0 (0.00)	0 (0.00)	—
Diabetes	10 (44.00)	3 (14.30)	—
Enlarged prostate	0 (0.00)	6 (29.00)	—
Epilepsy	1 (4.40)	0 (0.00)	—
Gastroparesis	1 (4.40)	0 (0.00)	—
Hypercholesterolemia	13 (56.52)	4 (19.04)	—
Hypertension	15 (65.22)	8 (38.10)	—
Hypothyroidism	3 (13.04)	0 (0.00)	—
Ischemic heart disease	0 (0.00)	2 (9.52)	—
Kidney failure	2 (9.00)	1 (5.00)	—
Macular degeneration	1 (4.40)	0 (0.00)	—
Mental health conditions	1 (4.40)	0 (0.00)	—
Pain	6 (26.10)	6 (29.00)	—
Sleep apnea	2 (9.00)	3 (14.30)	—
Stroke	4 (17.40)	3 (14.30)	—
Urinary retention	0 (0.00)	0 (0.00)	—
Other[e]	5 (21.74)	6 (29.00)	—

[a]Balance in the distribution of covariates between group 1 and 2 family health team sites was assessed using the Kruskal Wallis and Fisher exact test. Percentages may not be equal to 100% because of rounding.

[b]Family income before taxes in CAD $. US $1=CAD $1.3.

[c]University indicates individual has either completed a degree or is currently an undergraduate or graduate student.

[d]Missing data not applicable as there were not enough data per individual chronic illness to generate a meaningful *P* value.

[e]Mood disorders (anxiety or depression), multiple sclerosis, acute myocardial infarction, peripheral vascular disease, peripheral neuropathy, and osteoporosis.

Intervention Impact on Quality of Life and Self-management

Missing survey data were substantial, ranging between 14% and 91%, mainly because of nonresponse rather than attrition. There were 2 individuals who withdrew during the trial; therefore, the loss to follow-up was 4.5%.

Tables 3 and 4 present the descriptive statistics of the AQoL-4D and PAM scores from the sites at each data collection time point, where the *gray boxes* represent the control periods. Raw AQoL scores over time suggest that most patients (with the exception of those at site B) started and remained relatively healthy over the course of the study (with a notably wide SD).

After adjusting for the covariate of time in the model, patients with ePRO combined with usual care (mean 15.28, SD 18.60) demonstrated a nonsignificant decrease in quality of life (t_{24}=−1.20; *P*=.24) as compared with usual care only (mean 21.76, SD 2.17). With regard to patient engagement, ePRO combined with usual care (mean 66.5, SD 17.3) demonstrated a nonsignificant decrease in patient activation, t_{27}=−1.41; *P*=.17, as compared with usual care (mean 59.49, SD 9.60).

No patterns were evident when exploring descriptive trends in outcomes related to ePRO user intensity (eg, those who used the tool regularly versus those who rarely used or abandoned it all together). There were fewer completed follow-up surveys in the short term and nonuser groups.

Table 3. Mean (SD) of patient health-related quality of life at each discrete time point[a,b].

Calendar time	Group 1			Group 2		
	Site A (n=17)	Site D (n=5)	Site E (n=1)	Site B (n=13)	Site C (n=7)	Site F (n=1)
Baseline (January 2018)	*19.30 (10.10)*[c]	*22.22 (20.03)*	*6.00*	*10.61 (6.78)*	*28.00 (16.78)*	*17.00*
Period 1: April-July 2018	*20.94 (7.32)*	*28.47 (10.72)*	*6.00*	*11.11 (11.50)*	*31.00 (35.40)*	*6.00*
Period 2: July-October 2018	15.83 (7.64)	28.47 (22.00)	__[d]	*10.00 (5.74)*	*37.04 (23.62)*	*11.11*
Period 3: October 2018-January 2019	18.00 (10.00)	20.83 (25.53)	—	8.00 (8.19)	42.00 (35.40)	17.00
Period 4: January-April 2019	22.83 (12.94)	28.70 (13.13)	—	10.42 (9.00)	8.33 (8.00)[e]	22.22
Period 5: April-July 2019	11.11 (3.00)	36.11	—	—	—	—

[a]AQoL scoring 0 to 45 with 45 being the worst possible health.

[b]Mean (SD) quality-of-life scores could not always be calculated for each site and period because of missingness or lack of variability in the questionnaire responses.

[c]Italicization represents the usual care (control) period of the intervention.

[d]Missing data.

[e]For this site, there were only 2 respondents in periods 3 and 4. The two who responded in period 3 had a wide spread between scores (16.67 and 66.06), and the 2 respondents in period 4 were both lower overall (2.77 and 13.89).

Table 4. Mean (SD) of patient self-activation scores at each discrete time point during the trial[a]

Calendar time	Group 1			Group 2		
	Site A (n=17)	Site D (n=5)	Site E (n=1)	Site B (n=13)	Site C (n=7)	Site F (n=1)
Baseline (January 2018)	*60.42 (15.00)* [b]	*61.10 (9.00)*	*53.20*	*63.10 (15.00)*	*55.00 (11.30)*	*58.10*
Period 1: April-July 2018	*60.01 (10.59)*	*58.80 (8.26)*	*56.00*	*72.00 (19.12)*	*52.10 (1.60)*	*66.00*
Period 2: July-October 2018	70.00 (14.92)	63.10 (10.00)	56.00	*69.40 (19.04)*	*59.20 (9.00)*	*58.10*
Period 3: October 2018-January 2019	71.79 (20.31)	53.00 (15.00)	__[c]	68.00 (25.82)	56.30 (13.10)	56.00
Period 4: January-April 2019	68.57 (19.41)	63.00 (4.20)	—	98.00 (5.00)	56.00 (7.00)	61.00
Period 5: April-July 2019	73.57 (13.94)	48.90	—	76.93 (20.54)	—	66.00

[a]Mean (SD) patient self-activation scores could not always be calculated for each site and period because of missingness or lack of variability in the questionnaire responses.

[b]Italicization represents the usual care (control) period of the intervention.

[c]Missing data.

Mechanisms Likely Driving Outcomes: Selected Findings From Ethnographic Case Studies

Use log data revealed significant attrition on the tool for both long-and short-term user groups with 46% (21/46) of patients using the tool as intended, 15% (7/46) discontinued use after 3 months, and 36% (17/46) abandoned the app after initial training. Data from the ethnographic case studies are analyzed to provide insights into factors that may drive use and potentially influence outcomes.

Table 5 presents a summary of the data sources, and Multimedia Appendix 5 offers a summary of NPT constructs and analysis.

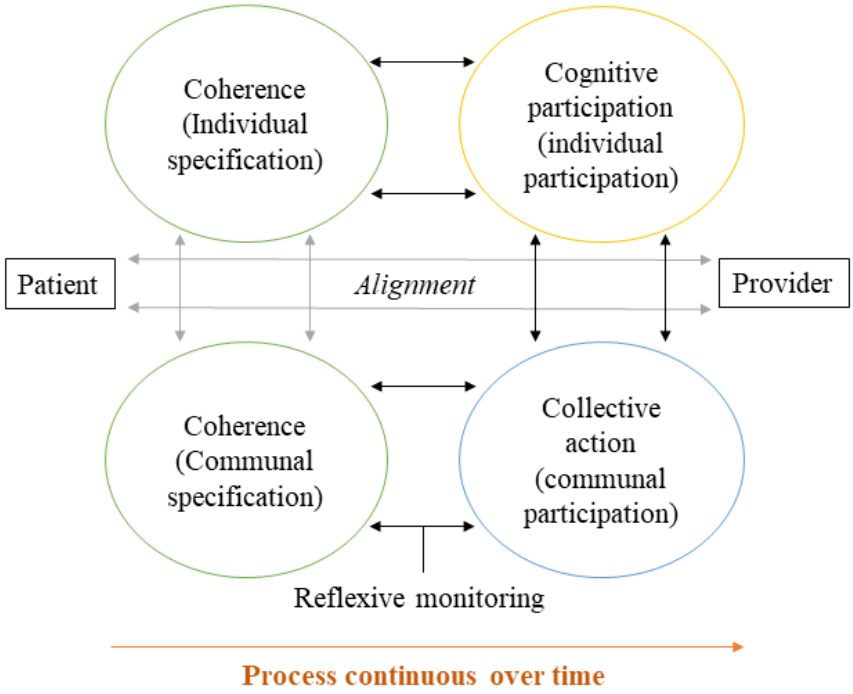
Table 5. Ethnographic data sources.

Case sites	Patient interviews (n=24)	Provider interviews (n=22)	Observations (n=21)
Site A	• 6 midterm • 5 end of project	• 6 midterm • 5 end of project	• 1 onboarding • 5 ad hoc
Site B	• 3 midterm • 3 end of project	• 4 midterm • 3 end of project	• 1 onboarding • 9 ad hoc
Site C	• 2 midterm • 2 end of project	• 2 midterm • 2 end of project	• 2 onboarding • 3 ad hoc

The Role of Coherence, Cognitive Participation, Collective Action, and Reflexive Monitoring

Consistent with findings from the exploratory trial, the meaningfulness of the ePRO tool to patients and providers had a significant influence on how and when it was used. Revealed in this analysis, is that the meaningfulness of the ePRO tool (its coherence to the participants) changed over time, and was reliant on (1) alignment to previously held notions of chronic disease management by providers and patients; (2) alignment to daily work and life activities (enabling cognitive participation); (3) strong relationships between patients and providers (enabling collective action); and (4) consistent positive assessments of the tool's utility (regular reflexive monitoring). An additional challenge is the interactional aspect of the ePRO tool, which means that both individual and collective coherence need to be aligned to the tool, as depicted in Figure 8.

Figure 8. Visual depiction of the normalization process of the electronic Patient-Reported Outcome tool.



The figure offers a simplified illustration of a complex ongoing process, highlighting 2 key drivers of adoption in this study. The first is the need for alignment between how individuals within a shared process understand that process (coherence) and then act on it (cognitive participation and collective action), depicted by gray arrows in the figure. Collective action (the actual use of the tool) proved to be highly influenced by this individual and shared coherence but differed depending on where users were in the process of GOC. For example, the data demonstrate that collective action occurred more toward the beginning of the intervention during goal-setting, when there was better alignment between individual and communal coherence of the intervention. This important time variable is indicated by the orange arrows. Second, the evaluation and assessment of the tool (reflexive monitoring) is continuous and interactive rather than a demonstration of normalization, as originally theorized, depicted in Figure 8 as black bidirectional arrows. As participants moved through participation and action in using ePRO, they consistently reflected on their individual and collective coherence, assessing whether it was worth continuing. Our data suggest that when alignment is high between individual- and group-level coherence, there is a greater likelihood of ongoing collective action; in this case, the use of the ePRO tool. This relationship is not currently depicted in the tool, as it will need to be tested in future studies.

Discussion

Participants and Study Implementation

Only 142 eligible patients of the total 176 patients were identified. The minimum recruitment numbers could not be reached owing to 3 challenges. First, some sites had few provider participants join the study. The usability study and exploratory trial suggested that providers who were just starting with the ePRO tool should manage a maximum of 5 patients at a time to reduce burden. The recruitment strategy required 6 to 8 providers to identify 5 to 10 patients each whom they could manage for the duration of the trial. As such, sites with fewer participants identified fewer patients to participate in the study (sites B and C in particular). Second, the requirement that patients *be ready* to engage in goal-setting proved to be a more significant barrier than at previous stages of ePRO testing. Practice EMRs identified many potential patient participants, but when reviewed by providers, few were identified as *ready.* Related to this point, the stepped-wedge design requires all participants to start an intervention at the same time. This rigid timing created an unanticipated third challenge.

The patient participants also likely represent a *healthier* group overall. Compared with similar patients in Canada, the United States, Australia, and the United Kingdom, patient participants had a lower number of reported chronic illnesses and a higher

level of reported education [64-67]. The AQoL scores of patients were aligned with previously published population norms [68]. Participants PAM standardized scores demonstrate slightly higher activation levels as compared with similar multimorbid populations, for instance, in a validation study of PAM that found a mean score of 56.6 (SD 12.9) [51].

Another important point to highlight is the relatively large number of registered dietitians and nurse practitioners who participated. One systematic review found several examples of dietitian-supported diabetes prevention programs [69], and nurse practitioners have been shown to successfully support digitally enabled chronic disease management programs in outpatient settings [70]. These examples, along with findings from this study, suggest an important role for allied health professionals in the implementation of digitally enabled health interventions for chronic disease populations in primary care settings.

Finally, the nature of the research process itself influenced trial implementation and outcome, as it conflicted with the natural process of delivering GOC. First, providers were exasperated by recruitment challenges, which resulted in delays in the trial start date. Second, while having providers manage few patients' reduced burden, it also meant providers had fewer opportunities to engage with the tool. As time went on, providers began to forget about the tool and why they valued it in the first place. Finally, the stepped-wedge required a time-bounded window for patient onboard. GOC, however, is a fluid approach in which goal-setting needs to occur at a point when patients are ready (as noted in the provider data around coherence). Providers expressed frustration that study parameters limited their ability to onboard patients later identified as individuals who could benefit from the tool.

Principal Findings

Recruitment challenges previously described resulted in the study being underpowered. As such, a conclusion regarding the effectiveness of the ePRO tool cannot be drawn. Analysis of the ethnographic data reveals interrelationships between use patterns, outcome trends, and patient and provider contexts to reveal the underlying mechanisms driving this complex intervention. Many patients and providers perceived the ePRO tool as valuable with the potential to improve engagement and healthy behaviors; however, over time, this excitement waned. Providers reverted to old ways of working, as did some patients. Waning engagement is not unique to digital health and has occurred in other behavior change interventions. Other patients for whom the tool was well-aligned to their values and aimed to manage their health demonstrated long-term adherence. For those *high users* whose coherence of the tool was tied to their interaction and relationship with their provider, they too began to fall away from the intervention as providers became less involved.

These findings uncover 2 tensions that have implications for digital health interventions for patients with complex care needs and multimorbidity in a primary care setting.

Challenge 1: Supporting Engagement in the Intervention Over Time

Engagement with an intervention is a well-documented challenge in primary care. Studies of medication adherence show similar ranges of adherence (40%-60%) [71,72] for chronic disease populations in primary care settings (ePRO adherence was 44%). *Nonadherence* reduces exposure and can lessen the effect of the intervention [73,74]. However, this lens suggests that it is the patient's fault for not doing as they are told, rather than placing a critical lens on the intervention itself. Perhaps a more useful lens is to consider engagement both "(1) the extent (eg, amount, frequency, duration, and depth) of use and (2) a subjective experience characterized by attention, interest, and affect" [75].

The ePRO tool experienced low retention rates typical of many mHealth interventions [76], which are connected to both use and subjective experience. The ethnographic findings suggest that subjective experience is linked to patient coherence and the meaningfulness of the tool. This finding is consistent with other studies that have shown that psychological factors such as motivation, expectations of the app, mental health, cognitive burden, and personal relevance will influence patient engagement [75]. Usability of the technology and tech savviness of users can often act as a barrier to ongoing use [77]. The usability analysis for this trial was too extensive to be included in this study. One key finding from the usability analysis presented in another paper, is that tech savviness and usability issues were moderated by the patient-provider relationship, in that patients with stronger regular connections to their providers were more likely to troubleshoot and work through technology challenges regardless of reported *savviness* [78].

Importantly, in this trial, app burnout occurred for patients *and* providers, for whom attrition was similarly linked to reduced use and subjective experience. Primary care providers have also demonstrated declining engagement with interventions over time, an issue identified in the literature as clinical inertia [79]. With continuous interventions, such as GOC, the ePRO study findings suggest that tapping into coherence consistently may improve engagement by both patients and providers.

Challenge 2: Meaningfulness for the Individual Versus the Group

Alignment of the ePRO tool to what was important and meaningful to patients and providers (coherence) was foundational. The study findings not only lend support for the importance of meaningfulness in technology [80] but also demonstrate how meaningfulness is constructed at both the individual and group levels, as suggested by NPT [60]. The disconnect between how providers and patients approached GOC is likely a contributor to the abandonment of the tool by those patients who sat somewhere between the strongly self-motivated super users and somewhat indifferent nonusers. For patients, GOC was a way to motivate and feel accountable for goals co-constructed with their providers. For many providers, however, GOC was an approach to support patient self-management, which did not require the same amount of ongoing connection and feedback expected from patients. This view is well-represented by the quote from the physical therapist

at site A, as shown in Multimedia Appendix 5, under the cognitive participation domain.

This approach to self-management suggests that a provider plays the role of a consultant, guiding patients through the management of their illnesses [81]. Although there is room for collaborative care and goal-setting in this model, the emphasis is on setting patients up to succeed and then sending them off. The qualitative data from this study suggest patients wanted more of a *coaching* approach with more touch points and interactions to maintain momentum, particularly for patients who started strong but then fizzled out. Theories of volition and self-regulation suggest that "feedback focused on the immediate benefits of behavior may be optimal during the early stages of behavior change" but can be reduced as individuals become more intrinsically motivated and confident [82]. What perhaps happened here is those patients who fizzled out were still at their *early stage* and, as such, required more engagement to keep moving forward. This finding suggests the need to better calibrate coherence when implementing digital health solutions with diverse user groups over time. Future studies should also probe how variation in the degree of goal specification found in this study may also influence patients at different stages of behavior change.

Strengths, Limitations, and Future Research

Similar to comparable studies of digital health interventions in primary care [83], both site and patient recruitment challenges were experienced. In addition, some values in the sample size calculation, such as attrition, were underestimated. Underpowering meant that all confounding variables collected via demographic baseline surveys and chart reviews could not be included in the modeling. In addition, some context data (notably participation in chronic disease management programs) may have changed over time and were only collected once at baseline. The smaller than anticipated sample size did allow for a more robust approach to ethnographic data collection, resulting in a rich, qualitative data set, which is a strength of the study. Future studies in primary care settings should consider both the setting context and the nature of the intervention being tested to better align trial methods to real-world implementation. More flexible adaptive trials or the application of an interrupted time series within the clusters may be more appropriate in these dynamic environments. Further exploration as to the reason why some providers were more successful in identifying eligible patients as compared with others in the study is another potential area of study to better understand this implementation challenge.

The findings may not be widely generalizable to older adults with complex needs, as patients in this study were generally healthier and more educated. However, the high proportion of complex older women living in rural environments in this project addresses a notable gap in the evidence on interventions for this population [84,85]. Relying on provider screening may have led to selection bias, which can reduce generalizability. However, as there is a lack of consensus on the definition of patient complexity, reliance on physician expertise and self-identification has been found to be a viable approach to identify this patient population [86].

Another important limitation is that this study lacks additional data on provider characteristics, such as age, years of experience, and employment status (eg, full-time or part-time). While a baseline demographic survey was deployed to all providers, few remitted these surveys despite multiple attempts to collect the data either via email, phone, or in person. While some key variables, such as comfort with technology, were collected via interviews, the other demographic variables would have aided in interpreting data and supported generalizability to other similar provider groups.

While this study offers a multimethod view of the effectiveness of the ePRO tool, the findings presented here focus on the major themes that emerged in the analysis. Further analyses will explore the interrelationships between NPT constructs and other context variables, in particular how these concepts relate to professional identities, organizational culture, and notions of how best to engage in chronic disease management. An important lesson from this trial is how the nature of GOC and chronic disease management in primary care settings is a fluid and complex process that is often unique to particular settings and provider-patient pairs. Highly adaptive trial designs, which allow the study to align to these contexts more closely, may have greater success in engagement for longer interventions.

Conclusions

Although this study is unable to provide a definitive answer to the effectiveness of the ePRO tool, it did generate novel insights regarding the implementation of digital health technologies in primary care settings. The findings demonstrate the critical role of coherence, or meaningfulness, of an intervention, and the great challenge of aligning coherence across diverse user groups over time. Future work in this area should pay careful attention to how chronic disease management, GOC, and self-management are understood and pursued when implementing digital health technologies to advance these models of care.

Acknowledgments

Funding support for this project was obtained from the Canadian Institutes of Health Research eHealth Innovation Partnership Program operating grant (FN-143559). The co-authors would like to acknowledge and thank the providers, administrators, and patients who participated in this study. The authors are grateful to each of them who gave their time, enthusiasm, and energy. The authors would also like to thank and acknowledge research support from Jasvinei Srinatharan, Tujuanna Austin, Karen Paik, Candis LePage, Janelle Gravesande, Zaki Hyatt-Shaw, and Parminder Hans for their coordination support throughout this study. They also acknowledge the contributions of collaborators Dr Cheryl Cott and Dr Renee Lyons, who offered early guidance in this work.

Conflicts of Interest

Funding for this study was obtained from a Canadian Federal Grant (Canadian Institutes of Health Research, FN-143559), with most co-authors' salaries being funded through their academic and scientific positions at their respective institutions. One co-author, SH, is the co-founder and co-owner of the technology company that hosts the electronic Patient-Reported Outcome (ePRO) tool (QoC Health Inc). While QoC Health Inc owns the platform on which the ePRO tool sits, the tool itself is the intellectual property of the lead author (CSG) in partnership with the Health System Performance Network at the University of Toronto, which is led by WPW (senior author). Any future aims to commercialize the ePRO tool would be done on the foundation of building a research-supporting not-for-profit that would seek to advance the development and evaluation of technologies that support care delivery for patients with complex care needs.

Multimedia Appendix 1
PRECIS-2 (Pragmatic Explanatory Continuum Indicator Summary) Wheel domain description.
[DOCX File , 20 KB - jmir_v23i12e29071_app1.docx]

Multimedia Appendix 2
Electronic Patient-Reported Outcome wireframes.
[PDF File (Adobe PDF File), 884 KB - jmir_v23i12e29071_app2.pdf]

Multimedia Appendix 3
Data collection schedule.
[DOCX File , 20 KB - jmir_v23i12e29071_app3.docx]

Multimedia Appendix 4
Sample interview guide questions.
[DOCX File , 17 KB - jmir_v23i12e29071_app4.docx]

Multimedia Appendix 5
Summary of how patients and providers understood, engaged with, and reflected on the adoption of the electronic Patient-Reported Outcome tool as aligned with Normalization Process Theory constructs.
[DOCX File , 28 KB - jmir_v23i12e29071_app5.docx]

Multimedia Appendix 6
Types of goals and tasks set during the trial.
[DOCX File , 24 KB - jmir_v23i12e29071_app6.docx]

References

1. 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]
2. Dobrow M. Caring for people with chronic conditions: a health system perspective. Int J Integr Care 2009 Mar 16;9(1):298. [doi: 10.5334/ijic.298]
3. Schaink AK, Kuluski K, Lyons RF, Fortin M, Jadad AR, Upshur R, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework. J Comorb 2012 Oct 10;2(1):1-9 [FREE Full text] [doi: 10.15256/joc.2012.2.15] [Medline: 29090137]
4. Bayliss E, Bosworth H, Noel P, Wolff J, Damush T, Mciver L. Supporting self-management for patients with complex medical needs: recommendations of a working group. Chronic Illn 2007 Jun 29;3(2):167-175. [doi: 10.1177/1742395307081501] [Medline: 18083671]
5. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
6. 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]
7. Berntsen G, Strisland F, Malm-Nicolaisen K, Smaradottir B, Fensli R, Røhne M. The evidence base for an ideal care pathway for frail multimorbid elderly: combined scoping and systematic intervention review. J Med Internet Res 2019 Apr 22;21(4):e12517 [FREE Full text] [doi: 10.2196/12517] [Medline: 31008706]
8. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]

9. Evangelista L, Steinhubl SR, Topol EJ. Digital health care for older adults. Lancet 2019 Apr;393(10180):1493. [doi: 10.1016/s0140-6736(19)30800-1]
10. Herzer KR, Pronovost PJ. Ensuring quality in the era of virtual care. J Am Med Assoc 2021 Feb 02;325(5):429-430. [doi: 10.1001/jama.2020.24955] [Medline: 33528544]
11. Chehade MJ, Yadav L, Jayatilaka A, Gill TK, Palmer E. Personal digital health hubs for multiple conditions. Bull World Health Organ 2020 Jun 02;98(8):569-575. [doi: 10.2471/blt.19.249136]
12. Sinnott C, Mc Hugh S, Browne J, Bradley C. GPs' perspectives on the management of patients with multimorbidity: systematic review and synthesis of qualitative research. BMJ Open 2013 Sep 13;3(9):e003610 [FREE Full text] [doi: 10.1136/bmjopen-2013-003610] [Medline: 24038011]
13. Tinetti ME, Naik AD, Dodson JA. Moving from disease-centered to patient goals-directed care for patients with multiple chronic conditions: patient value-based care. JAMA Cardiol 2016 Apr 01;1(1):9-10 [FREE Full text] [doi: 10.1001/jamacardio.2015.0248] [Medline: 27437646]
14. Reuben DB, Tinetti ME. Goal-oriented patient care — an alternative health outcomes paradigm. N Engl J Med 2012 Mar;366(9):777-779. [doi: 10.1056/nejmp1113631]
15. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition--multimorbidity. J Am Med Assoc 2012 Jun 20;307(23):2493-2494 [FREE Full text] [doi: 10.1001/jama.2012.5265] [Medline: 22797447]
16. World Health Organization. Regional Office for the Western Pacific. People-Centred Health Care : A Policy Framework. Manila: WHO Regional Office for the Western Pacific; 2007:1-20.
17. Coulter A, Entwistle VA, Eccles A, Ryan S, Shepperd S, Perera R. Personalised care planning for adults with chronic or long-term health conditions. Cochrane Database Syst Rev 2015 Mar 03;2015(3):CD010523 [FREE Full text] [doi: 10.1002/14651858.CD010523.pub2] [Medline: 25733495]
18. Ahmad N, Ellins J, Krelle H, Lawrie M. Person-centred care: from ideas to action. The Health Foundation. 2014. URL: https://www.health.org.uk/publications/person-centred-care-from-ideas-to-action [accessed 2021-10-09]
19. Mold JW, Blake GH, Becker LA. Goal-oriented medical care. Fam Med 1991 Jan;23(1):46-51. [Medline: 2001782]
20. Secunda K, Wirpsa MJ, Neely KJ, Szmuilowicz E, Wood GJ, Panozzo E, et al. Use and meaning of "Goals of Care" in the healthcare literature: a systematic review and qualitative discourse analysis. J Gen Intern Med 2020 May 21;35(5):1559-1566 [FREE Full text] [doi: 10.1007/s11606-019-05446-0] [Medline: 31637653]
21. Bisognano M, Schummers D. Flipping healthcare: an essay by Maureen Bisognano and Dan Schummers. Br Med J 2014 Oct 03;349(oct03 3):g5852. [doi: 10.1136/bmj.g5852] [Medline: 25280777]
22. Schellinger SE, Anderson EW, Frazer MS, Cain CL. Patient self-defined goals: essentials of person-centered care for serious illness. Am J Hosp Palliat Care 2018 Jan 23;35(1):159-165 [FREE Full text] [doi: 10.1177/1049909117699600] [Medline: 28330379]
23. Tinetti ME, Naik AD, Dindo L, Costello DM, Esterson J, Geda M, et al. Association of patient priorities-aligned decision-making with patient outcomes and ambulatory health care burden among older adults with multiple chronic conditions: a nonrandomized clinical trial. JAMA Intern Med 2019 Oct 07;179(12):1688 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4235] [Medline: 31589281]
24. Berntsen GK, Dalbakk M, Hurley JS, Bergmo T, Solbakken B, Spansvoll L, et al. Person-centred, integrated and pro-active care for multi-morbid elderly with advanced care needs: a propensity score-matched controlled trial. BMC Health Serv Res 2019 Oct 03;19(1):682 [FREE Full text] [doi: 10.1186/s12913-019-4397-2] [Medline: 31581947]
25. Dolovich L, Oliver D, Lamarche L, Thabane L, Valaitis R, Agarwal G, et al. Combining volunteers and primary care teamwork to support health goals and needs of older adults: a pragmatic randomized controlled trial. Can Med Asso J 2019 May 06;191(18):491-500 [FREE Full text] [doi: 10.1503/cmaj.181173] [Medline: 31061074]
26. Pirhonen L, Olofsson EH, Fors A, Ekman I, Bolin K. Effects of person-centred care on health outcomes-A randomized controlled trial in patients with acute coronary syndrome. Health Policy 2017 Feb;121(2):169-179. [doi: 10.1016/j.healthpol.2016.12.003] [Medline: 28062091]
27. Kinmonth AL, Woodcock A, Griffin S, Spiegal N, Campbell MJ. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. The Diabetes Care From Diagnosis Research Team. Br Med J 1998 Oct 31;317(7167):1202-1208 [FREE Full text] [doi: 10.1136/bmj.317.7167.1202] [Medline: 9794859]
28. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Jun 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]
29. Steele Gray C, Shaw J. From summative to developmental: incorporating design-thinking into evaluations of complex interventions. J Integrat Care 2019 Jun 20;27(3):241-248. [doi: 10.1108/jica-07-2018-0053]
30. Steele Gray C, Miller D, Kuluski K, Cott C. Tying ehealth tools to patient needs: exploring the use of eHealth for community-dwelling patients with complex chronic disease and disability. JMIR Res Protoc 2014 Nov 26;3(4):e67 [FREE Full text] [doi: 10.2196/resprot.3500] [Medline: 25428028]
31. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]

32. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
33. Steele Gray C, Gravesande J, Hans PK, Nie JX, Sharpe S, Loganathan M, et al. Using exploratory trials to identify relevant contexts and mechanisms in complex electronic health interventions: evaluating the electronic patient-reported outcome tool. JMIR Form Res 2019 Feb 27;3(1):e11950 [FREE Full text] [doi: 10.2196/11950] [Medline: 30810532]
34. Tsoukas H. Don't simplify, complexify: from disjunctive to conjunctive theorizing in organization and management studies. J Manag Stud 2016 Jul 11;54(2):132-153. [doi: 10.1111/joms.12219]
35. 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]
36. Patton MC. Utilization-Focused Evaluation. Thousand Oaks, CA: Sage Publications Inc; 2008:1-688.
37. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. Br Med J 2015 Feb 06;350(feb06 1):h391 [FREE Full text] [doi: 10.1136/bmj.h391] [Medline: 25662947]
38. Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? Int J Epidemiol 2020 Jun 01;49(3):1043-1052 [FREE Full text] [doi: 10.1093/ije/dyaa077] [Medline: 32386407]
39. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016 Aug 04;375(5):454-463. [doi: 10.1056/nejmra1510059]
40. Greenhalgh T, Swinglehurst D. Studying technology use as social practice: the untapped potential of ethnography. BMC Med 2011 Apr 27;9(1):45 [FREE Full text] [doi: 10.1186/1741-7015-9-45] [Medline: 21521535]
41. Gray CS. Getting to goals: using the Electronic Patient Reported Outcome (ePRO) mobile app to support complex patients in primary care settings. In: Proceedings of the Annual Canadian Association of Health Services and Policy Research (CAHSPR) Conference. 2017 Presented at: Annual Canadian Association of Health Services and Policy Research (CAHSPR) Conference; May 2017; Toronto, Ontario p. 177 URL: https://cahspr.ca/wp-content/uploads/2020/11/Book-of-Abstracts-CAHSPR-2017.pdf
42. Glazier RH, Redelmeier DA. Building the patient-centered medical home in Ontario. J Am Med Assoc 2010 Jun 02;303(21):2186-2187. [doi: 10.1001/jama.2010.753] [Medline: 20516421]
43. Focus on georgraphy series, 2016 census. Statistics Canada. 2016. URL: https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Index-eng.cfm [accessed 2020-09-10]
44. Upshur R, Wang L, Moineddin R, Nie J, Tracy C. The complexity score: towards a clinically-relevant, clinician-friendly measure of patient multi-morbidity. Int J Person Cent Med 2012;2(4):315 [FREE Full text]
45. Hawthorne G, Korn S, Richardson J. Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust N Z J Public Health 2013 Feb;37(1):7-16. [doi: 10.1111/1753-6405.12004] [Medline: 23379800]
46. Osborne R, Hawthorne G, Lew EA, Gray LC. Quality of life assessment in the community-dwelling elderly validation of the Assessment of Quality of Life (AQoL) instrument and comparison with the SF-36. J Clin Epidemiol 2003 Feb;56(2):138-147. [doi: 10.1016/s0895-4356(02)00601-7]
47. Partners Advancing Transitions in Healthcare (PATH): Northumberland. The Change Foundation. 2015. URL: https://nhh.ca/document/path-news-release-pdf [accessed 2021-10-21]
48. Resources. ePRO. URL: https://www.eprobridgepoint.com/resources [accessed 2021-10-24]
49. AQoL instruments. Assessment of Quality of Life. 2020. URL: http://aqol.com.au/index.php/aqolinstruments [accessed 2020-12-03]
50. Hibbard J, 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]
51. Skolasky R, Green AF, Scharfstein D, Boult C, Reider L, Wegener ST. Psychometric properties of the patient activation measure among multimorbid older adults. Health Serv Res 2011 Apr;46(2):457-478 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01210.x] [Medline: 21091470]
52. Barker I, Steventon A, Williamson R, Deeny SR. Self-management capability in patients with long-term conditions is associated with reduced healthcare utilisation across a whole health economy: cross-sectional analysis of electronic health records. BMJ Qual Saf 2018 Dec 23;27(12):989-999 [FREE Full text] [doi: 10.1136/bmjqs-2017-007635] [Medline: 30139822]
53. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]

54. Bradway M, Pfuhl G, Joakimsen R, Ribu L, Grøttland A, Årsand E. Analysing mHealth usage logs in RCTs: explaining participants' interactions with type 2 diabetes self-management tools. PLoS One 2018 Aug 30;13(8):e0203202 [FREE Full text] [doi: 10.1371/journal.pone.0203202] [Medline: 30161248]
55. Davey C, Hargreaves J, Thompson JA, Copas AJ, Beard E, Lewis JJ, et al. Analysis and reporting of stepped wedge randomised controlled trials: synthesis and critical appraisal of published studies, 2010 to 2014. Trials 2015 Aug 17;16(1):358 [FREE Full text] [doi: 10.1186/s13063-015-0838-3] [Medline: 26278667]
56. Eikelenboom N, van Lieshout J, Jacobs A, Verhulst F, Lacroix J, van Halteren A, et al. Effectiveness of personalised support for self-management in primary care: a cluster randomised controlled trial. Br J Gen Pract 2016 Apr 14;66(646):354-361. [doi: 10.3399/bjgp16x684985]
57. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
58. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g]
59. Sandelowski M. Telling stories: narrative approaches in qualitative research. Image J Nurs Sch 1991;23(3):161-166. [doi: 10.1111/j.1547-5069.1991.tb00662.x] [Medline: 1916857]
60. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4(1):29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
61. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
62. Cresswell JW, Poth CN. Qualitative Inqiury and Research Design: Choosing Among Five Approaches, 4th ed. Thousand Oaks, CA: Sage Publications Inc; 2017:1-488.
63. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):73-86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
64. Salisbury C, Man M, Bower P, Guthrie B, Chaplin K, Gaunt DM, et al. Management of multimorbidity using a patient-centred care model: a pragmatic cluster-randomised trial of the 3D approach. Lancet 2018 Jul;392(10141):41-50. [doi: 10.1016/s0140-6736(18)31308-4]
65. Tracy CS, Bell SH, Nickell LA, Charles J, Upshur RE. The IMPACT clinic: innovative model of interprofessional primary care for elderly patients with complex health care needs. Can Fam Physician 2013 Mar;59(3):148-155 [FREE Full text] [Medline: 23486816]
66. Ploeg J, Matthew-Maich N, Fraser K, Dufour S, McAiney C, Kaasalainen S, et al. Managing multiple chronic conditions in the community: a Canadian qualitative study of the experiences of older adults, family caregivers and healthcare providers. BMC Geriatr 2017 Jan 31;17(1):40 [FREE Full text] [doi: 10.1186/s12877-017-0431-6] [Medline: 28143412]
67. Harrison C, Britt H, Miller G, Henderson J. Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice. BMJ Open 2014 Jul 11;4(7):e004694 [FREE Full text] [doi: 10.1136/bmjopen-2013-004694] [Medline: 25015470]
68. Hawthorne G, Richardson J, Day N. Using the Assessment of Quality of Life (AQoL), version 1. Technical Report 12 - Centre for Health Program Evaluation, Australia. 2009. URL: https://www.aqol.com.au/papers/techreport12.pdf [accessed 2021-10-10]
69. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]
70. Gordon K, Steele Gray C, Dainty KN, DeLacy J, Ware P, Seto E. Exploring an innovative care model and telemonitoring for the management of patients with complex chronic needs: qualitative description study. JMIR Nurs 2020 Mar 6;3(1):e15691 [FREE Full text] [doi: 10.2196/15691] [Medline: 34345777]
71. Fernandez-Lazaro CI, García-González JM, Adams DP, Fernandez-Lazaro D, Mielgo-Ayuso J, Caballero-Garcia A, et al. Adherence to treatment and related factors among patients with chronic conditions in primary care: a cross-sectional study. BMC Fam Pract 2019 Sep 14;20(1):132 [FREE Full text] [doi: 10.1186/s12875-019-1019-3] [Medline: 31521114]
72. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/nejmra050100]
73. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
74. Manwaring JL, Bryson SW, Goldschmidt AB, Winzelberg AJ, Luce KH, Cunning D, et al. Do adherence variables predict outcome in an online program for the prevention of eating disorders? J Consult Clin Psychol 2008;76(2):341-346. [doi: 10.1037/0022-006x.76.2.341]

75. 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]
76. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018 Jul 26;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
77. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
78. Tahsin F, Tracy S, Chau E, Harvey S, Loganathan M, McKinstry B, et al. Exploring the relationship between the usability of a goal-oriented mobile health application and non-usage attrition in patients with multimorbidity: a blended data analysis approach. Digit Health 2021 Oct 05;7:205520762110455. [doi: 10.1177/20552076211045579]
79. Ziemer DC, Doyle JP, Barnes CS, Branch WT, Cook CB, El-Kebbi IM, et al. An intervention to overcome clinical inertia and improve diabetes mellitus control in a primary care setting: Improving Primary Care of African Americans with Diabetes (IPCAAD) 8. Arch Intern Med 2006 Mar 13;166(5):507-513. [doi: 10.1001/archinte.166.5.507] [Medline: 16534036]
80. Steele Gray C. Seeking meaningful innovation: lessons learned developing, evaluating, and implementing the electronic patient-reported outcome tool. J Med Internet Res 2020 Jul 29;22(7):e17987 [FREE Full text] [doi: 10.2196/17987] [Medline: 32723725]
81. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
82. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015 Jul 17;1:2055207615595335 [FREE Full text] [doi: 10.1177/2055207615595335] [Medline: 29942544]
83. 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]
84. Liu KA, Mager NA. Women's involvement in clinical trials: historical perspective and future implications. Pharm Pract (Granada) 2016 Mar 06;14(1):708. [doi: 10.18549/pharmpract.2016.01.708]
85. Virani S, Burke L, Remick SC, Abraham J. Barriers to recruitment of rural patients in cancer clinical trials. J Oncol Pract 2011 May;7(3):172-177. [doi: 10.1200/jop.2010.000158]
86. Grant RW, Ashburner JM, Hong CS, Hong CC, Chang Y, Barry MJ, et al. Defining patient complexity from the primary care physician's perspective: a cohort study. Ann Intern Med 2011 Dec 20;155(12):797-804. [doi: 10.7326/0003-4819-155-12-201112200-00001] [Medline: 22184686]

Abbreviations

AFHTO: Association of Family Health Teams of Ontario
AQoL-4D: Assessment of Quality of Life–4 Dimensions
CIHR: Canadian Institutes of Health Research
eHIPP: eHealth Innovation Partnership Program
EMR: electronic medical record
ePRO: electronic Patient-Reported Outcome
FHT: Family Health Team
GOC: goal-oriented care
GP: general practitioner
mHealth: mobile health
NP: nurse practitioner
NPT: Normalization Process Theory
PAM-13: Patient Activation Measure–13 item
RD: registered dietitian
RN: registered nurse
SMART: specific, measurable, attainable, realistic, timely

Edited by R Kukafka; submitted 24.03.21; peer-reviewed by M Khetani, MF Alwashmi, T Patel, A Garcia; comments to author 15.06.21; revised version received 19.07.21; accepted 12.09.21; published 02.12.21.

Please cite as:
Steele Gray C, Chau E, Tahsin F, Harvey S, Loganathan M, McKinstry B, Mercer SW, Nie JX, Palen TE, Ramsay T, Thavorn K, Upshur R, Wodchis WP
Assessing the Implementation and Effectiveness of the Electronic Patient-Reported Outcome Tool for Older Adults With Complex Care Needs: Mixed Methods Study
J Med Internet Res 2021;23(12):e29071
URL: https://www.jmir.org/2021/12/e29071
doi:10.2196/29071
PMID:34860675

Original Paper

Information Patients With Melanoma Spontaneously Report About Health-Related Quality of Life on Web-Based Forums: Case Study

Rachel R J Kalf[1,2], MSc; Diana M J Delnoij[2,3], PhD; Bettina Ryll[4], MD, PhD; Marcel L Bouvy[1], PhD; Wim G Goettsch[1,2], PhD

[1]Department of Pharmacoepidemiology and Clinical Pharmacology, University Utrecht, Utrecht, Netherlands

[2]National Health Care Institute, Diemen, Netherlands

[3]Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, Netherlands

[4]Melanoma Patient Network Europe, Uppsala, Sweden

Corresponding Author:
Marcel L Bouvy, PhD
Department of Pharmacoepidemiology and Clinical Pharmacology
University Utrecht
Universiteitsweg 99
Utrecht, 3584 CG
Netherlands
Phone: 31 302537324
Email: m.l.bouvy@uu.nl

Abstract

Background: There is a general agreement on the importance of health-related quality of life (HRQoL). This type of information is becoming increasingly important for the value assessment of health technology assessment agencies in evaluating the benefits of new health technologies, including medicines. However, HRQoL data are often limited, and additional sources that provide this type of information may be helpful.

Objective: We aim to identify the HRQoL topics important to patients with melanoma based on web-based discussions on public social media forums.

Methods: We identified 3 public web-based forums from the United States and the United Kingdom, namely the Melanoma Patient Information Page, the Melanoma International Forum, and MacMillan. Their posts were randomly selected and coded using qualitative methods until saturation was reached.

Results: Of the posts assessed, 36.7% (150/409) of posts on Melanoma International Forum, 45.1% (198/439) on MacMillan, and 35.4% (128/362) on Melanoma Patient Information Page focused on HRQoL. The 2 themes most frequently mentioned were *mental health* and *(un)certainty*. The themes were constructed based on underlying and more detailed codes. Codes related to *fear, worry and anxiety*, *uncertainty*, and *unfavorable effects* were the most-often discussed ones.

Conclusions: Web-based forums are a valuable source for identifying relevant HRQoL aspects in patients with a given disease. These aspects could be cross-referenced with existing tools and they might improve the content validity of patient-reported outcome measures, including HRQoL questionnaires. In addition, web-based forums may provide health technology assessment agencies with a more holistic understanding of the external aspects affecting patient HRQoL. These aspects might support the value assessment of new health technologies and could therefore help inform topic prioritization as well as the scoping phase before any value assessment.

(J Med Internet Res 2021;23(12):e27497) doi:10.2196/27497

KEYWORDS

reimbursement decision-making; QoL; health care; quality of life

Introduction

Background

Decisions on the reimbursement of innovative medicines in Europe are most prominently based on the recommendations of national health technology assessment (HTA) agencies. Conventionally, these HTA recommendations are prepared directly after market-authorization of medicines. The starting point for these HTAs is the assessment of (added) therapeutic value, also known as relative effectiveness assessment, and subsequently, cost-effectiveness assessments. In both, relative effectiveness assessments and cost-effectiveness assessments, outcome measures such as the overall survival rate, adverse events (AEs), and health-related quality of life (HRQoL) are considered.

From the perspective of patients, HRQoL is an important outcome measure because it can capture how disease and treatment affect a patient's quality of life [1]. This is especially of interest in diseases such as cancer, where medicines may increase overall survival rates but may cause considerable toxicity. Therefore, HRQoL intends to inform HTA agencies on the relevance and added value of new oncology treatments for patients, for instance, if the medicine improves HRQoL by halting the progression of the disease, or alternatively, decreases HRQoL if toxicity or AEs have a large impact on the patient's well-being.

Although the assessment of HRQoL is becoming increasingly important in different areas of health care, relevant HRQoL data are often unavailable. For instance, patients with severe disease seem less likely to complete HRQoL questionnaires compared with their healthier counterparts [2,3]. The use of complicated HRQoL instruments increases respondent burden and may also lead to lower completion rates. Furthermore, patients might not be motivated to complete HRQoL questionnaires in a research setting if tangible respondent benefits are not delivered. Overall, HRQoL data are currently only sparsely represented in HTA reports for oncological products. More specifically, only in a third of HTA assessments were HRQoL data used [4], leading to a low impact of HRQoL on HTA decision-making despite the general recognition of the importance of HRQoL for patients and society.

In addition to the limited availability of HRQoL data, current methods used to measure HRQoL may fail to truly capture what is most relevant to patients [5], which may result in incorrect overall interpretation. Therefore, there is a continuous search for sources that provide additional relevant information on HRQoL. Social media is a convenient and well-established communication source and therefore presents an obvious potential option. Patients often use social media to gather information on their health condition and treatment options, to share their experiences, and to find social support [6-8]. Previous research has also shown that social media may help identify HRQoL topics of importance to patients, prioritize the topics most relevant to patients, or help in the distribution of HRQoL questionnaires [9-12]. Melanoma is an area of oncology that has seen the rapid introduction of several classes of new therapeutics with new modes of action, increasing the likelihood of existing HRQoL tools failing to capture patient-relevant outcomes [13-15]. Concomitantly, several web-based patient forums for melanoma have been active.

Objectives

To evaluate the potential relevance of social media as a meaningful source of HRQoL information for HTA, we identified the HRQoL topics that are most important to patients with melanoma based on discussions from web-based forums. Following the logic that in an unsupervised setting, patients would bring up topics relevant to them rather than being triggered by, for instance, a questionnaire, we focused on the research question: Which HRQoL topics do patients with melanoma and their caregivers spontaneously discuss on the web?

Methods

Overview

For this study, we focused on public web-based forums that are publicly accessible to anyone, as opposed to private patient communities. These public web-based forums provide peer support for a range of medical conditions, allowing the patients to share their experiences and provide information [16-18]. In a previous study, we collected patient perspectives on HRQoL from private social media sources, including a private Facebook (Facebook, Inc) group for patients with melanoma [9]. Using a different type of social media in this study allows for comparisons between the different sources of social media regarding the HRQoL topics discussed.

Selection of Web-Based Forums

The public web-based forums were identified using 2 internet search engines, namely (1) Google (Google, Inc) and (2) Bing (Microsoft, Inc), which are currently the most popular search engines in the world [19,20]. Searches were conducted in English, with a combination of the search terms *melanoma*, *forum*, *message board*, and *discussion board*. Browser history was cleared before each search because the previous searches might influence the search findings. This forum search was conducted on 5 consecutive days starting June 4, 2019, to account for the websites being unavailable owing to maintenance issues. The search results from the first 2 pages shown by (1) Google and (2) Bing were extracted and assessed for eligibility, and any advertisements or images were excluded.

Forums were eligible for inclusion based on the following three criteria: (1) the website had been active for ≥5 years based on the publication dates of posts, (2) at least 2000 posts had been posted on the forum, and (3) ≥5 new posts had been posted in the past week. We identified 3 forums as eligible for inclusion: Melanoma Patients Information Page (MPIP), Melanoma International Forum (MIF), and MacMillan Cancer Support Online Community for Melanoma Patients. Each forum was informed of our intention to use their publicly available posts for research purposes via email. Both MPIP and MIF are forums based in the United States and MacMillan is based in the United Kingdom. MPIP and MIF focus solely on patients with melanoma, whereas MacMillan provides information to support patients with cancer in general, in addition to having 64

cancer-specific forums (eg, melanoma, Hodgkin lymphoma, pancreatic cancer, and unknown primary cancer) [21-23].

Data Extraction

No login was used to gain access to any of the posts extracted, nor was login required on any of the forums. Each forum thread was sorted by the date of the last post, after which all the threads were collected. A thread is defined as a collection of posts, with an initial post that introduces a specific topic and the subsequent replies posted by one or more members of the forum. We collected the complete threads from each forum using the R package *rvest* (R Core Team) in September 2019 [24]. We collected the following data: title of the thread, text from each post, username of each post, date and time of each post, and whether a post was the original post or a reply. Each username was given a user ID to ensure anonymity.

Data Analysis

We coded the posts using the coding scheme developed in our previous study [9], in which members of the Melanoma Patient Network Europe, an established patient network for patients with melanoma, caregivers, and advocates, were approached via its multiple social media channels to anonymously complete a 25-item web-based survey. In this survey, questions regarding sociodemographic and clinical characteristics and several open questions exploring patient and caregiver perspectives on HRQoL (eg, "What is HRQoL in melanoma for you?", "Name 3 things that deteriorate your/the melanoma patient's HRQoL today?") were posed. Two researchers independently performed inductive content analysis on the responses to the open-ended questions and assigned codes, and any discrepancies in coding were resolved by consensus. As these themes and codes may not have covered all the topics spontaneously discussed in the forums, we created new themes and codes as required. The following themes were added: *alone* and *coping*, and the code *guilt* was added to the theme *certainty*. In addition, we adjusted the coding scheme to be more concise.

We excluded the posts that did not focus on HRQoL or melanoma, provided advice or shared experience, asked a question or provided information, or offered support. We defined HRQoL as the patient's subjective perception of the impact of the disease and its treatment on the physical, psychological, and social aspects of daily life [25,26]. From each forum, a random sample of 100 posts was coded by 4 authors (RRJK, DMJD, WGG and MLB). Agreement regarding the inclusion and exclusion of posts between the coders was 74% for MPIP (RRJK and DMJD), 85% for MIF (RRJK and WGG), and 83% for MacMillan (RRJK and MLB); any disagreements were discussed and resolved by consensus. From this random sample, 44% (44/100) were included in this study from MPIP, 61% (61/100) from MIF, and 63% (63/100) from MacMillan. Subsequently, author RRJK continued coding the posts selected randomly from each forum until 100 posts which referred to HRQoL aspects were included. After this, the posts were coded in batches of 25 until saturation. We defined saturation as not being able to identify a new emerging theme in 2 consecutive batches of 25 posts [27]. Owing to the vast number of posts in each forum, we decided to code until saturation because this was sufficient to identify which HRQoL aspects were relevant to patients with melanoma. When author RRJK was unsure about a specific post or code, the issue was discussed and resolved by consensus among authors (RRJK, DMJD, MLB, and WGG). A total of 72.4% (262/362) posts for MPIP, 75.6% (309/409) for MIF, and 77.2% (339/439) for MacMillan were assessed solely by author RRJK. Coder drift was not assessed in this study, and therefore, poses a potential coding bias.

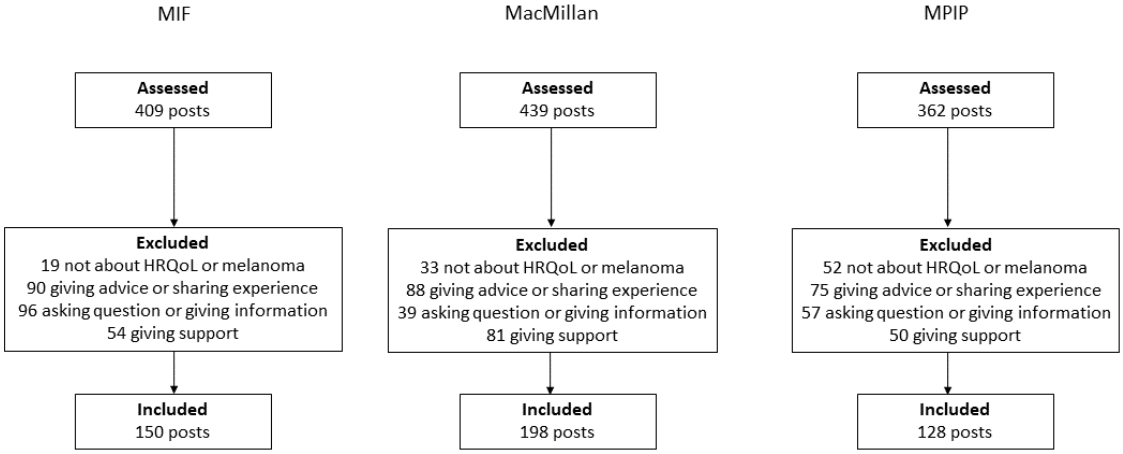
Covering all the posts assessed, we conducted an analysis of the number of threads and reply posts by each unique user to assess how often the same person initiated a thread or replied to an initial post. As a subanalysis, we assessed the subforums available on MIF in more detail. MIF provides separate forums for patients with melanoma with stage I and II, stage III and stage IV, as well as separate forums for newly diagnosed (ND) stage I and II patients and ND stage III and IV patients. This allowed us to evaluate which HRQoL topics were important for patients with melanoma at different disease stages. The results from analyzing the forum posts have been described qualitatively. This study was conducted in accordance with the Standards for Reporting Qualitative Research [28]. All data were collected, coded, stored, and analyzed using R version 3.4.4 (R Core Team) and NVivo (version 12; QSR International) [29,30].

Results

Overview

A total of 14,755, 6798, and 1671 threads were collected from MPIP, MIF, and MacMillan, respectively. This resulted in 88,261, 23,911, and 9551 original posts from MPIP, MIF, and MacMillan, respectively. A total of 409 posts from 189 unique users were assessed from MIF, as were 439 posts from 359 unique users from MacMillan and 362 posts from 243 unique users from MPIP (Figure 1). After the exclusion of irrelevant posts, 150 posts from 112 unique users, 198 posts from 164 unique users, and 128 posts from 96 unique users were included in our assessment from MIF, MacMillan, and MPIP, respectively. We determined how often the same user started a thread and posted a reply (Figure S1 in Multimedia Appendix 1). Some users started 1 thread but did not reply to any other post (46/189, 24.3% on MIF; 130/359, 36.2% on MacMillan; and 71/243, 29.2% on MPIP). Another group of users posted 1 reply, but did not start any threads (99/189, 52.4% MIF; 67/359, 18.7% MacMillan; and 52/243, 21.4% MPIP). Finally, a number of users started 1 thread and posted 1 reply (9/189, 4.8% MIF; 69/359, 19.2% MacMillan; and 32/243, 13.2% MPIP). Overall, 95.8% (181/189) of the users on MIF, 95% (341/359) on MacMillan, and 92.6% (225/243) on MPIP posted ≤5 posts (either as the initial thread or reply post). Only a few users in each forum contributed to a greater extent. Only 2 major outliers can be identified: 1 on MPIP, where 1 user started 45 threads and posted 26 replies, and 1 on MIF, where 1 user started 20 threads and posted 74 replies.

Figure 1. Overview of inclusion and exclusion criteria for forum posts. HRQoL: health-related quality of life; MIF: Melanoma International Forum; MPIP: Melanoma Patients' Information Page.



On all 3 forums, the 2 most often identified themes were *mental health* and *certainty* (Table 1). More than half of the posts mentioned aspects related to *mental health* (85/150, 56.7% MIF; 126/198, 63.6% MacMillan; and 69/128, 53.9% MPIP), and at least a third of the posts mentioned information relevant to *certainty* (63/150, 42% MIF; 80/198, 40.4% MacMillan; and 40/128, 31.3% MPIP). Other often-mentioned themes were *health care communication* (32/150, 21.3%) and *unfavorable effects* (28/150, 18.7%) on MIF, *health care access* (43/198, 21.7%) and *unfavorable effects* (27/198, 13.6%) on MacMillan, and *health care access* (20/128, 15.6%) and *unfavorable effects* (21/128, 16.4%) on MPIP.

Table 1. Total number and percentage of posts mentioning a specific theme on each forum (N=476).

Theme	Total posts per forum, n (%)		
	MIF[a] (n=150)	MacMillan (n=198)	MPIP[b] (n=128)
Mental health	85 (56.6)	126 (63.6)	69 (53.9)
Certainty	63 (42)	80 (40.4)	40 (31.2)
Health care communication	32 (21.3)	21 (10.6)	12 (9.4)
Unfavorable effects	28 (18.6)	27 (13.6)	21 (16.4)
Health care access	16 (10.6)	43 (21.7)	20 (15.6)
Health care general	17 (11.3)	23 (11.6)	18 (14.1)
Disease status	16 (10.6)	5 (2.5)	9 (7)
Support	15 (10)	26 (13.1)	12 (9.4)
Coping	14 (9.3)	22 (11.1)	4 (3.1)
Social life	14 (9.3)	19 (9.6)	17 (13.3)
Health general	13 (8.7)	7 (3.5)	6 (4.7)
Physical health	9 (6)	11 (5.6)	8 (6.3)
Treatment	9 (6)	4 (2)	14 (10.9)
Happiness	8 (5.3)	7 (3.5)	6 (4.7)
Alone	1 (0.7)	3 (1.5)	1 (0.8)

[a]MIF: Melanoma International Forum.

[b]MPIP: Melanoma Patients Information Page.

Each theme was constructed from underlying, more detailed codes. Table S1 in Multimedia Appendix 1 shows the codes used for each forum and provides excerpts from posts to provide examples for each code. This provides insight into the construct of each code and displays in more detail which HRQoL aspects the patients spontaneously discussed on the web. Examples of the most-often discussed codes (Table 2) are given below.

Table 2. Total number and percentage of posts mentioning a specific code on each forum (N=476).

Theme and code	Total posts per forum, n (%)		
	MIF[a] (n=150)	MacMillan (n=198)	MPIP[b] (n=128)
Mental health			
Fear, worry, and anxiety[c]	58 (38.8)	78 (39.4)	46 (35.9)
Positive mood	11 (7.3)	17 (8.6)	10 (7.8)
Mental health[d]	3 (2)	16 (8.1)	2 (1.6)
No anxiety or relieve	5 (3.3)	7 (3.5)	4 (3.1)
Stress	6 (4)	4 (2)	4 (3.1)
Not to worry	2 (1.3)	2 (1)	3 (2.3)
Depression	0 (0)	2 (1)	0 (0)
Certainty			
Uncertainty	38 (25.3)	46 (23.2)	20 (15.6)
Hope	13 (8.7)	18 (9.1)	13 (10.2)
Confusion	5 (3.3)	11 (5.6)	4 (3.1)
Guilt[d]	4 (2.7)	5 (2.5)	1 (0.8)
Confident	1 (0.7)	0 (0)	2 (1.6)
Control	2 (1.3)	0 (0)	0 (0)
Health care communication			
Lack of information	14 (9.3)	10 (5.1)	4 (3.1)
Informed decision-making	11 (7.3)	5 (2.5)	4 (3.1)
Good information	6 (4)	1 (0.5)	3 (2.3)
Counselling	1 (0.7)	2 (1)	1 (0.8)
Access to information	0 (0)	3 (1.5)	0 (0)
Unfavorable effect			
Unfavorable effects	25 (16.7)	24 (12.1)	19 (14.8)
No unfavorable effects	3 (2)	3 (1.5)	2 (1.6)
Health care access			
Waiting time	4 (2.7)	29 (14.6)	10 (7.8)
Finances	6 (4)	7 (3.5)	5 (3.9)
Access medicines	5 (3.3)	4 (2)	1 (0.8)
Access care	1 (0.7)	3 (1.5)	4 (3.1)
Health care general			
Good care or good doctors[c]	14 (9.3)	10 (5.1)	14 (10.9)
Bad care or bad doctors	3 (2)	13 (6.6)	4 (3.1)
Disease status			
No spreading	5 (3.3)	4 (2)	1 (0.8)
No evidence of disease	7 (4.7)	0 (0)	5 (3.9)
Progression	2 (1.3)	1 (0.5)	3 (2.3)
Metastasis	2 (1.3)	0 (0)	0 (0)
Support			
Support[d]	14 (9.3)	14 (7.1)	9 (7)
Ignorance	1 (0.7)	6 (3)	1 (0.8)

Theme and code	Total posts per forum, n (%)		
	MIF[a] (n=150)	MacMillan (n=198)	MPIP[b] (n=128)
Lack of support	0 (0)	6 (3)	2 (1.6)
Coping[d]	14 (9.3)	22 (11.1)	4 (3.1)
Social life			
Patient network	12 (8)	11 (5.6)	15 (11.7)
Work	2 (1.3)	3 (1.5)	1 (0.8)
Friends	0 (0)	2 (1)	1 (0.8)
Family	0 (0)	3 (1.5)	0 (0)
General health			
Pain	9 (6)	3 (1.5)	3 (2.3)
Diet and appetite	3 (2)	1 (0.5)	3 (2.3)
Good health	1 (0.7)	1 (0.5)	0 (0)
Pain free	0 (0)	2 (1)	0 (0)
Physical health			
Fatigue	5 (3.3)	7 (3.5)	5 (3.9)
Good physically	1 (0.7)	2 (1)	1 (0.8)
Pregnancy[d]	1 (0.7)	1 (0.5)	2 (1.6)
Exercise	2 (1.3)	1 (0.5)	0 (0)
Treatment			
Randomized controlled trials	5 (3.3)	1 (0.5)	6 (4.7)
Good medicines	1 (0.7)	3 (1.5)	7 (5.5)
Drug effectiveness	3 (2)	0 (0)	1 (0.8)
Happiness			
Enjoy life	4 (2.7)	5 (2.5)	1 (0.8)
Normal life	3 (2)	2 (1)	5 (3.9)
Capability	1 (0.7)	0 (0)	0 (0)
Alone[d]	1 (0.7)	3 (1.5)	1 (0.8)

[a]MIF: Melanoma International Forum.

[b]MPIP: Melanoma Patients Information Page.

[c]Codes combined as compared with coding scheme used in previous study.

[d]New codes added to the original coding scheme used in previous study [9].

Fear, Worry, and Anxiety

On all 3 forums, the code relating to *fear, worry, anxiety* was most often discussed (Table 2; Table S1 in Multimedia Appendix 1). More specifically, on all forums, users talked about being obsessed over moles and being scared about their diagnosis. Other aspects mentioned included, but were not limited to, being anxious about the results (MIF and MPIP), worrying about recurrences (MIF), and the consequences of stopping treatment (MIF and MPIP).

Uncertainty

The second most frequently discussed topic on MIF, MacMillan, and MPIP was *uncertainty* (Table 2). Users were uncertain about many different aspects, such as whether they had made the right decision, whether the medicines would work, if the diagnosis was correct, and how bad the AEs would be (Table S1 in Multimedia Appendix 1). For example, one user said "[...] any suggestions on [...] how not to worry endlessly about the 'what ifs'."

Unfavorable Effects

On MPIP, MacMillan, and MIF the topic *unfavorable effects* was also discussed commonly (Table 2). This focused on the AEs, the complications and the symptoms that the patients experienced. One specific AE that was most frequently mentioned on MIF and MacMillan was lymphedema (Table S1 in Multimedia Appendix 1). Users also discussed solutions to the AEs and the complications they were experiencing, such as those from the medicines they were prescribed (MIF,

MacMillan, MPIP). Not only were the intolerable AEs, complications, and symptoms discussed, but also those that were manageable. Discussions reflected the different degrees of AE presentation experienced by patients, from manageable to intolerable. For example, 1 user mentioned the following "not the end of the world itching and rash, but it is very maddening and crazy making.", while another indicated "[...] has a terrible rash on his face head and back. We can LIVE with the rash."

Waiting Time and Coping

On MacMillan, next to *unfavorable effects*, both *waiting time* and *coping* were often mentioned (Table 2). Coping was also a topic discussed on the other 2 forums (Table S1 in Multimedia Appendix 1), although it seemed to be discussed to a lesser extent. Users discussed how they coped, for example, with their diagnoses (MIF and MacMillan), with the AEs (MIF and MacMillan), and with their lives in the new normal (MIF and MPIP). Some users indicated how difficult it was to cope with their diagnosis and how they went through denial before being able to accept the seriousness of it all (MacMillan). The long waiting time for appointments and results were also mentioned on all forums (Table S1 in Multimedia Appendix 1). Users expressed this as: *feels like waiting for eternity*, and *the waiting game being the worst*. However, some users on MacMillan also indicated that the waiting time was not as long as anticipated.

Hope

Hope was also a code mentioned in all forums. On MIF, a user expressed the following: "I'm getting to the point where I'm believing I could be ok again!" Users also expressed their hope of having scans that showed tumor shrinkage (MacMillan and MPIP) and their hope for medicines that would work (MacMillan and MPIP).

Health Care

Members shared experiences related not only to their health, but also to their experiences with health care, including access to health care, lack of information, and making informed decisions (Table S1 in Multimedia Appendix 1). On all forums, users talked about good and bad experiences with their health care. For example, one user posted:

> *I was seen by a new (?), certainly very young doctor who had obviously not read my notes as he had no idea that I was on the Avastin trial. In fact he didn't even know what the trial was and even asked me to spell the drug's name for him!!!! Obviously a very well read young man in his specialist field, not!*

However, good experiences with health care were also shared, such as by this user:

> *Had my first PET this week since stage 4 dx, and met with Onc the same day to go over results. She hadn't looked at them yet when we met, so I was pretty nervous. She could tell and just told me these are the first scans and the only bad results would be if there are any new mets that had popped up in kidney, lungs, or any other organs. She said she would be happy with no change, or even if things only grew by a little.*

The subanalysis of MIF subforums (data not shown) showed that fear, worry and anxiety was discussed on all subforums, but most often by patients with stage I or II, with 55.0% (33/60; including ND) of the posts mentioning this topic. Uncertainty was discussed on all subforums to approximately the same extent (17.6% (6/34) - 32.3% (10/31) of the posts discussed this topic). The topic unfavorable effects was more often discussed by stage III and IV patients (25.6% (22/86) including ND) than by stage I and II patients (5.0% (3/60) including ND). ND patients discussed coping more often than patients who were not posting on the ND subforums (17.2% (11/64) vs 3.7% (3/82), respectively).

Discussion

Principal Findings

In this study, we showed that patients with melanoma and their caregivers discussed many different topics related to HRQoL on public web-based forums. Topics related to *fear, worry and anxiety*, *uncertainty*, and *unfavorable effects* were most often discussed. With respect to *fear, worry and anxiety*, some users discussed their worries regarding their moles and diagnosis, which may be most important to patients in the earlier stages of melanoma. Other users discussed aspects related to their fear of recurrence or the consequences of stopping treatment, which may be more relevant to patients in the later stages of the disease. Of note, a caveat of social media is the incomplete information on user characteristics, making it infeasible to determine the disease stage for each user. Many users also discussed aspects related to *uncertainty*. However, this covered different aspects ranging from uncertainty regarding AEs and the effectiveness of the medicines to uncertainty about their diagnosis. Finally, with respect to discussions on unfavorable effects, users shared their experiences with AEs and complications, as well as their solutions.

It is important to realize that the type of social media used may affect the results of a study like ours because social media may be public (anyone may gain access to posts without signing in) or private (where an account is needed before users may gain access to posts). In public sources, users may be less inclined to share personal experiences as compared with private social media sources [31,32]. Previously, we had assessed which HRQoL topics were most important to patients with melanoma and their caregivers on private social media by posting a survey on the private social media channels of Melanoma Patient Network Europe [9]. It was shown that *family* and having a *normal life* were the 2 most important HRQoL topics for patients with melanoma. In this study, patients with melanoma most often discussed topics related to *fear, worry, anxiety, uncertainty*, and *unfavorable effects*. This difference may be because in the previous paper, we actively inquired about the HRQoL aspects most important to patients with melanoma, guiding them through a survey, whereas in the current paper, we merely listened to the topics that patients with melanoma discussed with each other [33,34], the latter being a much more inductive approach.

Another aspect that may influence our study results is the overrepresentation of a specific group of users, such as patients

with a specific stage of disease or their caregivers discussing the topics most important to them and subsequently driving our results. We previously showed using private social media that patients with melanoma with a different stage of the disease find other HRQoL aspects important, as do caregivers [9]. In this study, we confirmed this as our subanalysis of the MIF subforums suggested that different HRQoL topics seemed important to patients with melanoma in different disease stages. Subsequently, melanoma-specific HRQoL questionnaires may benefit from taking these differences into account.

Previous research has shown that disease-specific HRQoL questionnaires do not fully represent what patients find important in HRQoL [9,12,35]. For example, the wording in the questionnaires may be different from how patients describe HRQoL aspects; some topics may seem less relevant to patients and some topics may not be included in the HRQoL questionnaires [9,12,35]. Therefore, we evaluated whether melanoma-specific HRQoL questionnaires represented topics discussed by patients with melanoma on web-based forums. In both the Functional Assessment of Cancer Therapy-Melanoma (FACT-M) and European Organization for Research and Treatment of Cancer (EORTC) QLQ-MEL38, some questions related to the theme *mental health* are present [36,37]. In FACT-M, these questions seem to focus on worrying, losing hope, being sad, and feeling nervous. Although in EORTC QLQ-MEL38, they seem to focus mainly on worrying, including worrying about unfavorable effects. In contrast, web-based discussions seem to focus more on the (overwhelming) fear and anxiety of patients with melanoma. Regarding *uncertainty*, only EORTC QLQ-MEL38 poses one question *Have you felt able to plan for the future?* However, in web-based discussions, other aspects of uncertainty seem to be more important to patients. Several questions related to *unfavorable effects* are posed in both FACT-M and EORTC QLQ-MEL38, including some questions related to lymphedema. This seems to correspond to the web-based discussions among patients with melanoma. Other themes that were often discussed on the web included *health care communication* and *health care access*. It seems that only EORTC QLQ-MEL38 has questions focusing on these themes. Although these melanoma-specific HRQoL questionnaires have been developed with great care, these findings raise questions about the extent to which these questionnaires cover aspects most pertinent to patients. Therefore, HRQoL questionnaires may benefit from ensuring that topics correspond more to patient experiences, such as including more questions on uncertainty.

It is important to note that although aspects related to AEs may be important for reimbursement decision-making, aspects related to uncertainty and coping are less relevant. However, considering the high psychological burden in the early stages of melanoma, which contrasts with the seemingly benign overall survival outcomes, some topics may become increasingly relevant for HTA as melanoma therapies move from the advanced setting into earlier stages of the disease. These insights highlight the importance as well as the burden that these topics present for patients with melanoma across all disease stages, in addition to disease-specific concerns. Health care systems, therefore, should be aware that topics such as health care communication and access to services can critically impact the HRQoL of patients, irrespective of the given treatment.

Limitations

This study has several limitations. First, we focused on web-based forums, whereas other public, social media channels might provide a different type of insight (eg, Twitter, public Facebook groups, or blogs). However, not every social media channel is appropriate for gathering insights on a specific topic. For example, information on AEs is not readily available on Facebook or Twitter [38]. Second, identifying the disease stage for each patient was difficult but has been proven to be relevant as our earlier analysis of stage-specific forums has shown. This could possibly be overcome to a certain degree by using more automated methods of data analysis. In addition, validating authenticity (eg, verifying whether users actually have the disease they discuss) on the web is difficult [11,39,40]. Third, selection bias may be an issue because the patient population present on web-based forums may be different from the patient population that is not using web-based forums. For example, patients using social media are conventionally better educated [40,41], more likely to be female [39,42], and may have a different symptom experience [43]. Finally, web-based forums may update their terms of use at any given time. At the time of collecting the posts, all 3 forums were of public nature and logging in to gain access to the posts was not necessary. However, MIF has changed this and now requires a login to gain access to posts.

Strengths

One of the strengths of this study was the coding of 100 posts from each forum by 2 authors to ensure validity. Analyzing qualitative data can be subjective; therefore, agreement among multiple authors when assigning codes is important. In addition, any uncertainties in the posts that were coded until saturation was reached were discussed and resolved by consensus among the 4 authors to further ensure validity. Another strength was determining how often users posted an initial post and a reply post to assess whether one or more users could possibly drive our results. A total of 94.4% (747/791)of users posted only a few posts (≤5 posts) on the forums, suggesting that our results were not driven by one or more users. It seems highly unlikely that the 2 outliers in MIF and MPIP would drive the results, considering the number of posts assessed.

Implications and Conclusions

Patient involvement is becoming increasingly important in HTA, which is especially appreciated during the scoping phase of HTA and for HTA topic prioritization [44,45]. The scoping phase is conducted at the beginning of an HTA assessment, where the technology under assessment, the reference or comparator technology, the relevant study population, and the relevant outcome measures regarding the effectiveness and safety of the technology under assessment are identified [46,47]. In an ideal situation, several stakeholders, including clinicians and patients, should be involved during this scoping phase. The input from patients or their representatives is, for example, important to choose outcome measures that matter to patients. However, the involvement of patients or representatives may

be limited [48-51]. Social media may allow inputs from a wide group of patients, and may thus provide robust insight into patient experiences. For example, in the scoping phase, social media may be informative in determining which outcome measures would be most important to measure. Social media may not only prove useful in HTA but may also inform health care professionals in their understanding of patient experiences and what is important to patients regarding their treatment and health care [52]. In addition, issues relevant to patients and which they deal with on a regular basis may be uncovered and could lead to identifying issues that might have otherwise gone unrecognized [53]. In addition, in regulatory decision-making, information from social media may help determine which AEs greatly affect HRQoL, are most debilitating to patients, and which AEs are acceptable to patients [39,54,55]. Therefore, social media may be informative for several stakeholders with varying goals. However, this source of information still needs to become part of the regular data extraction practices of stakeholders. Therefore, clear guidelines are needed for the ethical use of social media data, the limitations that are involved, and the purposes for which this information could be used.

To conclude, it is important to realize that web-based forums are a valuable source to cross-reference the relevance of existing tools and help identify gaps in existing procedures. Social media may contribute to improving the content validity of patient-reported outcome measures, including HRQoL measures. More specifically, current melanoma HRQoL questionnaires may potentially improve patient relevance by adding more items related to fear, worry, anxiety and uncertainty. Social media is a readily available source that can provide fast inputs from patients with both rare and common diseases. It can be used passively to listen to what patients discuss on the web and to actively distribute questionnaires. In addition, information extracted from social media may support an evidence ecosystem, where existing evidence is used by several stakeholders for different goals. This information source may contribute to a more holistic understanding of the patient's perspective and highlight external factors affecting patient HRQoL. Social media may specifically provide insights for HTA decision-making during the prioritization of topics as well as during the scoping phase conducted before the value assessment of a new health technology.

Acknowledgments

This study was funded by the Dutch National Health Care Institute.

Conflicts of Interest

MPNE (Melanoma Patient Network Europe) is a volunteer-based network whose activities are funded by balanced support by the following companies: Amgen, BMS, Delcath, Idera, Incyte, Immunocore, MSD, Merck Serono, Novartis, Pierre Fabre, and Roche. Support never includes editorial rights, influence on MPNE's program nor activities. In addition, MPNE is funded by 3 Horizon2020 grants (UMCURE, Share4Rare and iTobos). BR received personal consultancy fees for work in patient affairs from Amgen, AstraZeneca, BMS, Celgene, Clinigen, IQVIA, MSD, Novartis, Pfizer, and Roche.

Multimedia Appendix 1
An overview of the codes per forum including examples illustrating each code, and the number of threads and reply posts each unique user posted on each forum.
[PDF File (Adobe PDF File), 167 KB - jmir_v23i12e27497_app1.pdf]

References

1. Kleijnen S, Leonardo Alves T, Meijboom K, Lipska I, De Boer A, Leufkens HG, et al. The impact of quality-of-life data in relative effectiveness assessments of new anti-cancer drugs in European countries. Qual Life Res 2017 Sep;26(9):2479-2488 [FREE Full text] [doi: 10.1007/s11136-017-1574-9] [Medline: 28401419]
2. Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. BMC Cancer 2016 Feb 18;16:122 [FREE Full text] [doi: 10.1186/s12885-016-2152-1] [Medline: 26892541]
3. Grady KL, Jones PG, Cristian-Andrei A, Naftel DC, Myers S, Dew MA, et al. Causes and consequences of missing health-related quality of life assessments in patients who undergo mechanical circulatory support implantation: insights from INTERMACS (Interagency Registry for Mechanically Assisted Circulatory Support). Circ Cardiovasc Qual Outcomes 2017 Dec;10(12):e003268 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.116.003268] [Medline: 29246883]
4. Kleijnen S, Lipska I, Leonardo Alves T, Meijboom K, Elsada A, Vervölgyi V, et al. Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. Ann Oncol 2016 Sep;27(9):1768-1775 [FREE Full text] [doi: 10.1093/annonc/mdw233] [Medline: 27329251]
5. Carr A, Higginson IJ. Are quality of life measures patient centred? BMJ 2001 Jun 02;322(7298):1357-1360 [FREE Full text] [doi: 10.1136/bmj.322.7298.1357] [Medline: 11387189]
6. Sillence E. Giving and receiving peer advice in an online breast cancer support group. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):480-485. [doi: 10.1089/cyber.2013.1512] [Medline: 23659724]

7. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007 Oct;68(2):173-178. [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
8. 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 18;13(10):e0205855 [FREE Full text] [doi: 10.1371/journal.pone.0205855] [Medline: 30335805]
9. Makady A, Kalf RR, Ryll B, Spurrier G, de Boer A, Hillege H, et al; GetReal Workpackage 1. Social media as a tool for assessing patient perspectives on quality of life in metastatic melanoma: a feasibility study. Health Qual Life Outcomes 2018 Nov 29;16(1):222 [FREE Full text] [doi: 10.1186/s12955-018-1047-z] [Medline: 30497502]
10. Kalf RR, Makady A, Ten Ham RM, Meijboom K, Goettsch WG, IMI-GetReal Workpackage 1. Use of social media in the assessment of relative effectiveness: explorative review with examples from oncology. JMIR Cancer 2018 Jun 08;4(1):e11 [FREE Full text] [doi: 10.2196/cancer.7952] [Medline: 29884607]
11. Zaid T, Burzawa J, Basen-Engquist K, Bodurka DC, Ramondetta LM, Brown J, et al. Use of social media to conduct a cross-sectional epidemiologic and quality of life survey of patients with neuroendocrine carcinoma of the cervix: a feasibility study. Gynecol Oncol 2014 Jan;132(1):149-153 [FREE Full text] [doi: 10.1016/j.ygyno.2013.10.015] [Medline: 24145111]
12. Cotté FE, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, et al. Exploring the health-related quality of life of patients treated with immune checkpoint inhibitors: social media study. J Med Internet Res 2020 Sep 11;22(9):e19694 [FREE Full text] [doi: 10.2196/19694] [Medline: 32915159]
13. Franklin C, Livingstone E, Roesch A, Schilling B, Schadendorf D. Immunotherapy in melanoma: recent advances and future directions. Eur J Surg Oncol 2017 Mar;43(3):604-611. [doi: 10.1016/j.ejso.2016.07.145] [Medline: 27769635]
14. Larocca CA, LeBoeuf NR, Silk AW, Kaufman HL. An update on the role of Talimogene Laherparepvec (T-VEC) in the treatment of melanoma: best practices and future directions. Am J Clin Dermatol 2020 Dec;21(6):821-832. [doi: 10.1007/s40257-020-00554-8] [Medline: 32767272]
15. Domingues B, Lopes J, Soares P, Populo H. Melanoma treatment in review. Immunotargets Ther 2018 Jun 7;7:35-49 [FREE Full text] [doi: 10.2147/itt.s134842]
16. Brady E, Segar J, Sanders C. Accessing support and empowerment online: the experiences of individuals with diabetes. Health Expect 2017 Oct;20(5):1088-1095 [FREE Full text] [doi: 10.1111/hex.12552] [Medline: 28718928]
17. Verberne S, Batenburg A, Sanders R, van Eenbergen M, Das E, Lambooij MS. Analyzing empowerment processes among cancer patients in an online community: a text mining approach. JMIR Cancer 2019 Apr 17;5(1):e9887 [FREE Full text] [doi: 10.2196/cancer.9887] [Medline: 30994468]
18. Gupta T, Schapira L. Online communities as sources of peer support for people living with cancer: a commentary. J Oncol Pract 2018 Oct 18:JOP1800261 (forthcoming). [doi: 10.1200/JOP.18.00261] [Medline: 30335558]
19. Bhasin H. Top 15 most popular search engines of the world. Marketing 91. 2019. URL: https://www.marketing91.com/search-engines-of-the-world/ [accessed 2019-09-01]
20. Law T. Meet the top 10 search engines in the world in 2021. Oberlo. 2020. URL: https://www.oberlo.com/blog/top-search-engines-world [accessed 2021-11-22]
21. Melanoma International Foundation. URL: https://forum.melanomainternational.org/ [accessed 2021-11-22]
22. Melanoma Patients Information Page. Melanoma Research Foundation. URL: https://melanoma.org/legacy/find-support/patient-community/mpip-melanoma-patients-information-page [accessed 2021-11-22]
23. Online community. MacMillan Cancer Support. URL: https://community.macmillan.org.uk/g [accessed 2021-11-22]
24. Wickham H. Rvest: Easily Harvest (Scrape) Web Pages. R package version 0.3. 2021. URL: https://cran.r-project.org/web/packages/rvest/rvest.pdf [accessed 2021-11-22]
25. Endpoints used for relative effectiveness assessment: health-related quality of life and utility measures. European Network for Health Technology Assessment. 2013. URL: https://tinyurl.com/2p9ab24f [accessed 2021-11-22]
26. Reflection paper on the use of patient reported outcome (PRO) measures in oncology studies. Medicines and Healthcare Products Regulatory Agency. 2014. URL: https://www.ema.europa.eu/en/documents/presentation/presentation-reflection-paper-use-patient-reported-outcome-pro-measures-oncology-studies-daniel_en.pdf [accessed 2021-11-22]
27. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
28. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
29. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: https://www.R-project.org/ [accessed 2021-11-22]
30. NVivo Qualitative Data Analysis Software. NVivo. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-11-22]

31. Årsand E, Bradway M, Gabarron E. What are diabetes patients versus health care personnel discussing on social media? J Diabetes Sci Technol 2019 Mar;13(2):198-205 [FREE Full text] [doi: 10.1177/1932296818821649] [Medline: 30636445]
32. Apperson A, Stellefson M, Paige S, Chaney B, Chaney J, Wang M, et al. Facebook groups on chronic obstructive pulmonary disease: social media content analysis. Int J Environ Res Public Health 2019 Oct 09;16(20):3789 [FREE Full text] [doi: 10.3390/ijerph16203789] [Medline: 31600907]
33. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291. [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
34. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. J Clin Epidemiol 2008 Oct;61(10):1049-1055. [doi: 10.1016/j.jclinepi.2007.11.012] [Medline: 18538997]
35. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
36. Functional Assessment of Cancer Therapy – Melanoma. Fact-M. URL: https://www.facit.org/measures/FACT-M [accessed 2021-11-22]
37. European Organization for Research and Treatment of Cancer core Quality of Life Questionnaire for Melanoma. URL: https://qol.eortc.org/questionnaires/ [accessed 2021-11-22]
38. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol 2015 Oct;80(4):878-888 [FREE Full text] [doi: 10.1111/bcp.12746] [Medline: 26271492]
39. van der Heijden L, Piner SR, van de Sande MA. Pigmented villonodular synovitis: a crowdsourcing study of two hundred and seventy two patients. Int Orthop 2016 Dec;40(12):2459-2468. [doi: 10.1007/s00264-016-3208-1] [Medline: 27169531]
40. Marshall SA, Yang CC, Ping Q, Zhao M, Avis NE, Ip EH. Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. Qual Life Res 2016 Mar;25(3):547-557 [FREE Full text] [doi: 10.1007/s11136-015-1156-7] [Medline: 26476836]
41. Beusterien K, Tsay S, Gholizadeh S, Su Y. Real-world experience with colorectal cancer chemotherapies: patient web forum analysis. Ecancermedicalscience 2013 Oct 10;7:361 [FREE Full text] [doi: 10.3332/ecancer.2013.361] [Medline: 24143155]
42. McCarrier KP, Bull S, Fleming S, Simacek K, Wicks P, Cella D, et al. Concept elicitation within patient-powered research networks: a feasibility study in chronic lymphocytic leukemia. Value Health 2016 Jan;19(1):42-52 [FREE Full text] [doi: 10.1016/j.jval.2015.10.013] [Medline: 26797235]
43. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013 Mar;22(3):256-262 [FREE Full text] [doi: 10.1002/pds.3365] [Medline: 23322591]
44. Gagnon M, Candas B, Desmartis M, Gagnon J, La Roche D, Rhainds M, et al. Involving patient in the early stages of health technology assessment (HTA): a study protocol. BMC Health Serv Res 2014 Jun 20;14:273 [FREE Full text] [doi: 10.1186/1472-6963-14-273] [Medline: 24950739]
45. Facey K. Patient involvement in HTA: what added value? Pharm Policy Law 2011;13:245-251. [doi: 10.3233/ppl-2011-0329]
46. Pakketadvies in de praktijk - Wikken en wegen voor een rechtvaardig pakket. National Health Care Institute Netherlands. 2017. URL: https://www.zorginstituutnederland.nl/publicaties/rapport/2017/09/06/rapport-pakketadvies-in-de-praktijk-wikken-en-wegen-voor-een-rechtvaardig-pakket [accessed 2021-10-23]
47. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence. 2013. URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [accessed 2021-11-22]
48. Menon D, Stafinski T. Role of patient and public participation in health technology assessment and coverage decisions. Expert Rev Pharmacoecon Outcomes Res 2011 Feb;11(1):75-89. [doi: 10.1586/erp.10.82] [Medline: 21351860]
49. Bridges JF, Jones C. Patient-based health technology assessment: a vision of the future. Int J Technol Assess Health Care 2007;23(1):30-35. [doi: 10.1017/S0266462307051549] [Medline: 17234014]
50. Hicks N, Hawken N, Arvin-Berod C, Toumi M. A literature review of Patient Advocacy Group (Pag) involvement in HTA. Value Health 2014 Nov;17(7):A438 [FREE Full text] [doi: 10.1016/j.jval.2014.08.1139] [Medline: 27201166]
51. Hailey D, Werkö S, Bakri R, Cameron A, Göhlen B, Myles S, et al. Involvement of consumers in health technology assessment activities by INAHTA agencies. Int J Technol Assess Health Care 2013 Jan;29(1):79-83. [doi: 10.1017/S026646231200075X] [Medline: 23217279]
52. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med 2019 Jan 16;17(1):11 [FREE Full text] [doi: 10.1186/s12916-018-1247-8] [Medline: 30646913]
53. Wu H, Fang H, Stanhope SJ. Exploiting online discussions to discover unrecognized drug side effects. Methods Inf Med 2013;52(2):152-159. [doi: 10.3414/ME12-02-0004] [Medline: 23450374]
54. 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]

55. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: lessons from IMI WEB-RADR. Drug Saf 2019 Dec;42(12):1393-1407 [FREE Full text] [doi: 10.1007/s40264-019-00858-7] [Medline: 31446567]

Abbreviations

AE: adverse event
EORTC: European Organization for Research and Treatment of Cancer
FACT-M: Functional Assessment of Cancer Therapy-Melanoma
HRQoL: health-related quality of life
HTA: health technology assessment
MIF: Melanoma International Forum
MPIP: Melanoma Patients' Information Page
ND: newly diagnosed

Edited by R Kukafka; submitted 27.01.21; peer-reviewed by N Lau, L Guo; comments to author 02.04.21; revised version received 27.08.21; accepted 25.09.21; published 07.12.21.

Please cite as:
Kalf RRJ, Delnoij DMJ, Ryll B, Bouvy ML, Goettsch WG
Information Patients With Melanoma Spontaneously Report About Health-Related Quality of Life on Web-Based Forums: Case Study
J Med Internet Res 2021;23(12):e27497
URL: https://www.jmir.org/2021/12/e27497
doi:10.2196/27497
PMID:34878994

Original Paper

Population Preferences for Performance and Explainability of Artificial Intelligence in Health Care: Choice-Based Conjoint Survey

Thomas Ploug[1], PhD; Anna Sundby[1], PhD; Thomas B Moeslund[2], PhD; Søren Holm[3], PhD

[1]Department of Communication and Psychology, Aalborg University, Copenhagen, Denmark

[2]Visual Analysis and Perception Lab, Aalborg University, Aalborg, Denmark

[3]Centre for Social Ethics and Policy, University of Manchester, Manchester, United Kingdom

Corresponding Author:
Thomas Ploug, PhD
Department of Communication and Psychology
Aalborg University
A C Meyers Vænge 15
Copenhagen, 2450
Denmark
Phone: 45 99402533
Email: ploug@hum.aau.dk

Abstract

Background: Certain types of artificial intelligence (AI), that is, deep learning models, can outperform health care professionals in particular domains. Such models hold considerable promise for improved diagnostics, treatment, and prevention, as well as more cost-efficient health care. They are, however, opaque in the sense that their exact reasoning cannot be fully explicated. Different stakeholders have emphasized the importance of the transparency/explainability of AI decision making. Transparency/explainability may come at the cost of performance. There is need for a public policy regulating the use of AI in health care that balances the societal interests in high performance as well as in transparency/explainability. A public policy should consider the wider public's interests in such features of AI.

Objective: This study elicited the public's preferences for the performance and explainability of AI decision making in health care and determined whether these preferences depend on respondent characteristics, including trust in health and technology and fears and hopes regarding AI.

Methods: We conducted a choice-based conjoint survey of public preferences for attributes of AI decision making in health care in a representative sample of the adult Danish population. Initial focus group interviews yielded 6 attributes playing a role in the respondents' views on the use of AI decision support in health care: (1) type of AI decision, (2) level of explanation, (3) performance/accuracy, (4) responsibility for the final decision, (5) possibility of discrimination, and (6) severity of the disease to which the AI is applied. In total, 100 unique choice sets were developed using fractional factorial design. In a 12-task survey, respondents were asked about their preference for AI system use in hospitals in relation to 3 different scenarios.

Results: Of the 1678 potential respondents, 1027 (61.2%) participated. The respondents consider the physician having the final responsibility for treatment decisions the most important attribute, with 46.8% of the total weight of attributes, followed by explainability of the decision (27.3%) and whether the system has been tested for discrimination (14.8%). Other factors, such as gender, age, level of education, whether respondents live rurally or in towns, respondents' trust in health and technology, and respondents' fears and hopes regarding AI, do not play a significant role in the majority of cases.

Conclusions: The 3 factors that are most important to the public are, in descending order of importance, (1) that physicians are ultimately responsible for diagnostics and treatment planning, (2) that the AI decision support is explainable, and (3) that the AI system has been tested for discrimination. Public policy on AI system use in health care should give priority to such AI system use and ensure that patients are provided with information.

(J Med Internet Res 2021;23(12):e26611) doi:10.2196/26611

KEYWORDS

artificial Intelligence; performance; transparency; explainability; population preferences; public policy

Introduction

Recent developments in artificial intelligence (AI) hold considerable promise for promoting individual health and well-being, and societal flourishing. Taking medical imaging as an example, a recent review showed that although the diagnostic performance of deep learning models is generally equivalent to that of health care professionals, it may outperform such professionals in particular cases [1]. Better diagnosis and early detection may not only enable better treatment but also lead to more cost-effective public spending. However, the performance of AI models comes at a cost. The most successful deep learning models are opaque. The complexity of such models implies (1) that many aspects of the decision-making procedure cannot be fully explicated and scrutinized and (2) that the exact reasoning cannot be replicated step-by-step in real time [2]. Across all types of AI, it seems that at the moment, the better-performing models (eg, deep learning models) are the most opaque [2-8].

The opacity of a medical AI system may concern different groups in the clinical setting. Thus, there are partly dissociable transparency and explainability problems in relation to health care professionals and in relation to patients. If an AI system and its outputs can be made transparent to health care professionals, then patients can potentially rely on the trust they have in those professionals, even if the patients themselves do not understand the system. However, if there is also a significant transparency and explainability problem in relation to health care professionals, both patients and professionals end up in a situation in which the only possible reliance on trust will be on trust in the system. This paper primarily investigates this as seen from a patient perspective.

Opacity is undesirable. The transparency and explainability of AI decision making to the patient are important for psychological, scientific, ethical, and democratic reasons. Psychologically, transparency may increase the understanding of the AI process and may make it easier to cope with AI decisions significantly impacting individual lives (eg, receiving a life-changing diagnosis). Scientifically, transparency may provide insights into hitherto unknown correlations constitutive or suggestive of causal mechanisms. Ethically, transparency may provide a basis for individual self-protection against biased and discriminatory decisions, decisions based on violations of privacy, decisions subjecting individuals to unreasonable risks of harm, etc. Democratically, transparency may unveil the inner workings of the technology used in specific settings by the state or other powerful actors to exercise power over citizens, and thus, it may empower citizens to hold decision makers accountable through the institutions of democracy.

The importance of transparency and explainability of AI is widely recognized. Researchers and research institutions, public committees, and expert groups, as well as private companies, have in recent years issued guidelines for responsible use of AI, emphasizing the value of transparency. A recent systematic review found 84 such guidelines [9]. The authors showed a remarkable global convergence on the importance of transparency, but they also exhibited a significant variation in what transparency is taken to be and requires. The importance of transparency is recognized in the European Union General Data Protection Regulation (GDPR), which in Articles 13 and 14 stipulates that if data subjects are profiled, they have a right to "meaningful information about the logic involved" [10]. A right to "meaningful information" is, however, rather vague. It may be interpreted minimally as simply requiring abstract and generic information about AI involvement in decision making along the lines of "this decision was partly based on recommendations made by an automated computer system." It may also be interpreted maximally as requiring access to all aspects of the AI decision making and the ability to reproduce each and every (significant) step in the decision making. As noted above, a maximal interpretation would entail that some of the best-performing systems of AI cannot satisfy the requirement of transparency.

The need for transparency is also recognized by researchers and developers of AI systems, and "explainable AI" is an active research field, and it is probably unlikely that any system would be implemented in health care without some work having been done estimating the importance of features such as gender, age, and ethnicity on the outputs of the system. This is, however, still not full transparent or explainable.

This paper proceeds from the assumption that there is a real dilemma here. Maximal transparency of AI systems may come at the cost of system performance and vice versa. This is not a conceptually necessary dilemma. New AI architectures may be invented that satisfy all relevant criteria of transparency and explainability and at the same time perform better than current architectures. However, until that happens, there is a balance to be struck between transparency and system performance. We believe that an adequate requirement of transparency should consider individuals' interests and preferences for performance and transparency. We therefore studied the relative importance to citizens of these and other aspects of AI decision making in health care in the Danish population by performing a conjoint analysis survey.

Methods

Initial Focus Groups

In this study, 2 focus group interviews were conducted with 5-6 participants in each drawn from Kantar Gallup´s Danish consumer panel. The participants were a cross section of the public and ranged from age 27 to 75 years. Both interviews were conducted in September 2019, and each interview lasted about 2 hours. The participants were briefly introduced to AI and were subsequently presented with 2 scenarios revolving around the use of AI for decision making in health care. They were asked to discuss each of the scenarios and in particular (1) the importance of being provided with explanations of the AI decision making and (2) the trade-off between the accuracy and performance of AI decision making and being provided with explanations of the decision making. The groups were asked, for instance, how important it is to explain how AI reaches decisions, even if the ability to do so will make the AI decisions less accurate.

All interviews were audio-recorded and transcribed verbatim. The interviews in combination with the literature offered information that was used to identify 6 aspects that play a role in the participants' views on the use of AI in health care: (1) type of AI decision (ie, whether it is used for diagnostics or treatment planning), (2) the level of explanation available, (3) performance and accuracy, (4) responsibility for the final decision, (5) possibility of bias or discrimination, and (6) severity of the disease or condition to which the AI is applied.

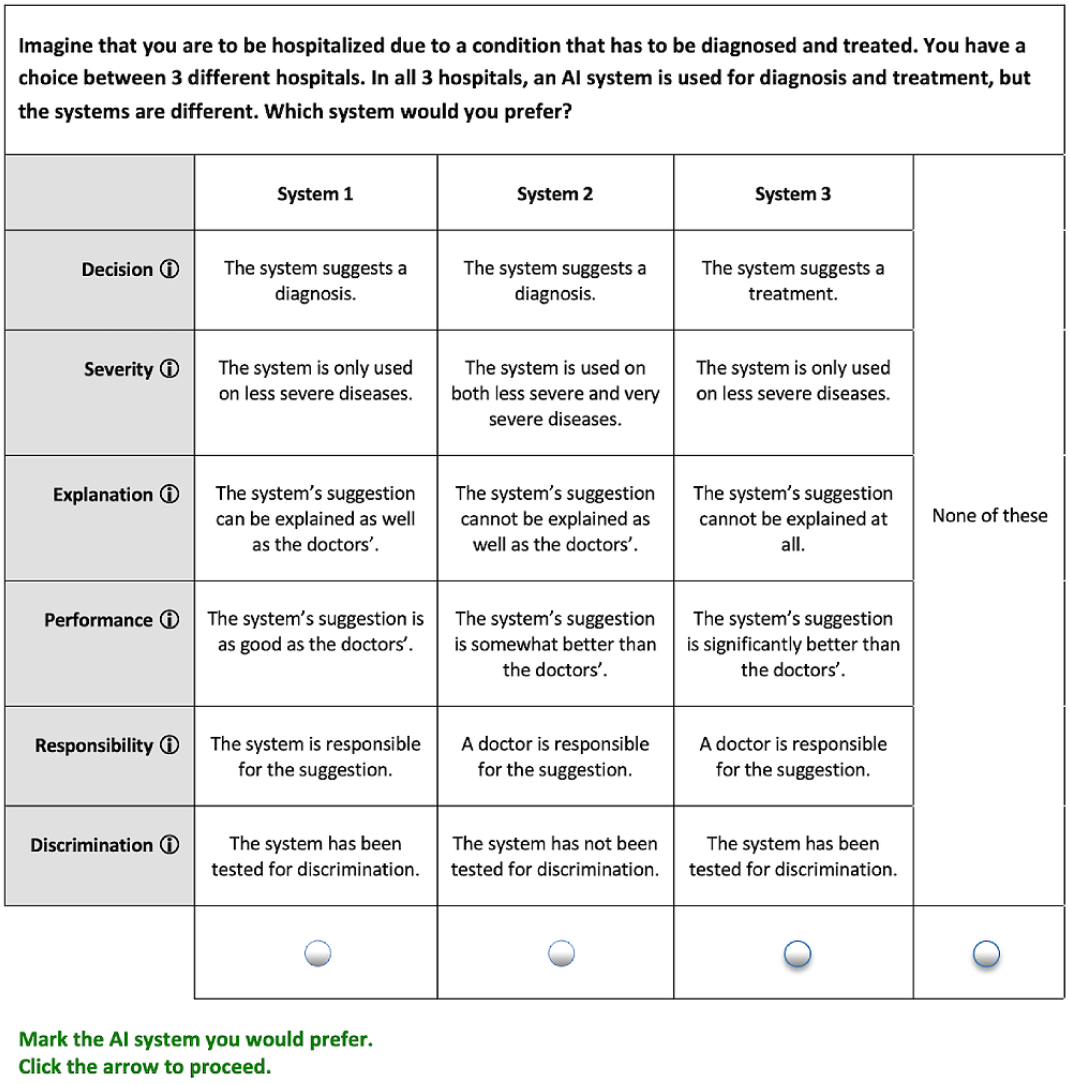
Design of Survey

Conjoint analysis is a discrete choice survey methodology. Respondents are asked to make a choice between 2 or more different options, where each option is described in terms of a number of predefined attributes, each with a number of levels (Figure 1). Given a sufficient number of choices per respondent, it is then possible to statistically estimate the importance of each attribute and level for the choice in terms of part-worth utilities [11].

The 6 aspects mentioned above were chosen as attributes for the conjoint analysis survey. For each attribute, a number of levels were developed based on the literature. In setting the lowest level for performance, it was assumed that any AI system introduced in health care would be known to perform at least as well as a trained health care professional (see also the Discussion section later).

The choice sets were generated using the complete enumeration method in the Sawtooth SSI Web (version 7.0.30) module [12]. The complete enumeration method generates conjoint designs conforming to the principles of (1) minimal overlap of attribute levels within a single choice task, (2) level balance across the set of choice tasks presented to each respondent, and (3) orthogonality (ie, the levels of different attributes are chosen independently). In total, 100 unique sets of conjoint choice questionnaires were generated, each of which was presented to an approximately equal number of respondents. Each set of conjoint choice questionnaires contained 12 choice tasks, where each respondent was asked to choose 1 of 3 options or a "None of these" option.

Figure 1. An example of a choice task with 3 concepts. AI: artificial intelligence.



The choice situation was described as follows:

> *This study is about artificial intelligence (AI). AI is a way of getting digital technologies to solve complex tasks. There is, for instance, AI in self-driving cars, search engines on the web, or the voice assistants in mobile phones. AI is often based on large data sources. You can, for instance, train an AI system to make diagnoses of cataract or melanoma by showing it many different pictures of eyes or skin with moles. You can also train AI systems to make suggestions about treatment of a disease (eg, suggestions for medication). AI systems have shown themselves to be quite good at making diagnoses and suggestions about treatment. However, they also sometimes make errors or differentiate unjustly between patients. In addition, it can be difficult to explain how a diagnosis or a suggestion for treatment has been derived. In what follows, you will therefore be asked about what you would prioritize if an AI system was used in the health care sector to make a diagnosis or suggest a treatment for you. On the next 12-15 pages, you will be shown 3 different scenarios that all involve the use of AI for diagnosis and treatment. On each page, you should choose the scenario that you think is best/you prefer. If none of the scenarios look good to you, there is an option "None of these."*

In addition to the conjoint analysis survey, respondents were asked about demographic data, chronic illness, and recent contact with the health care system; questions about trust in the health care system; and questions about fear and hope in relation to AI in general [13].

Sample

A stratified sample of 1678 potential participants was drawn from Kantar Gallup's Danish consumer panel of 53,000 active members. The sample was designed to be representative of the adult Danish population. Emails were sent to the potential participants, inviting them to participate in the study. After 3, 11, and 29 days, nonresponders (ie, those who had not completed the survey or who had not visited the website hosting the survey) were reminded by email. After 6, 22, and 31 days, nonresponders were contacted by SMS.

Of the 1678 potential respondents contacted, 1441 opened the link to the questionnaire, 1027 completed it fully, and 414 completed it partially. The analysis was based on the 1027 complete answers, giving a response rate of 61.2% (1027/1678). A sample efficiency analysis calculating the overall concordance between the respondents and the desired sample characteristics was performed, considering obtained and desired numbers in relation to gender, age, geographical region, and level of education (sum of squares=33.78, *df*=785, efficiency=88.33%).

Statistical Analysis

The analysis of the conjoint analysis survey deriving the part-worth utilities of the attributes and levels was performed by Kantar Gallup. Part-worth utilities were estimated using Sawtooth CBC/HB (version 5.5.3) to perform a hierarchical Bayes method estimation, running 190,000 burn-in iterations and 10,000 draws per respondent. A detailed description of the hierarchical Bayes method and its implementation can be found in Ref. [14].

The subsequent statistical analysis was performed using IBM SPSS Statistics 25. Demographic data were tabulated and univariate relationships between utilities and respondent characteristics analyzed using ANOVA with Bonferroni correction. Ad hoc trust, fear, and hope scales were formed as simple summative scales from the trust, fear, and hope questions and validated by Cronbach α. All 3 scales had acceptable α values of .74-.79. Univariate relationships between utilities and the 3 scales were analyzed using correlation analysis with Bonferroni correction.

Results

Major Findings

A total of 521 of 1027 (50.7%) respondents were men and 506 (49.3%) women. The average age was 50.3 years (SD 18.1). Of the 1027 respondents, 375 (36.5%) indicated that they had a chronic illness, 830 (80.8%) had visited their general practitioner (GP) at least once during the past year, and 146 (14.2%) had been inpatients in a hospital during the past year. The highest educational level was school or high school for 197 of 1027 (19.2%) respondents, further education for 446 of 1027 (43.4%) respondents, and university or university college for 384 of 1027 (37.4%) respondents.

The part-worth utilities of the attributes and levels are presented in Table 1, the responses to questions about trust in health care and technology, and fear and hope in relation to AI in general in Table 2, and the relationship between respondent characteristics and utilities of attributes in Table 3.

Table 1. Importance of attributes and part-worth utilities of levels.

Attribute	Importance (%)	Level (part-worth utility)
Type	3.0	• Diagnostics (0.123) • Treatment planning (–0.123)
Explanation	27.3	• Equally explainable as physician's decision (1.106) • Not as explainable as physician's decision (–0.270) • No explanation available (–0.836)
Performance	6.6	• System decision significantly better than physician's (0.267) • System decision somewhat better than physician's (0.052) • System decision equally good as physician's (–0.319)
Responsibility	46.8	• Physician responsible for decision (1.900) • System responsible for decision (–1.900)
Discrimination	14.8	• System tested for biased decisions (0.602) • System not been tested for biased decisions (–0.602)
Severity of disease	1.5	• System use only when less severe disease (0.060) • System use both when less severe and when very severe disease (–0.060)

Table 2. Respondent trust and opinions about AI[a] (N=1027).

Opinion	None/not at all, n (%)	Very little, n (%)	Little, n (%)	Some, n (%)	A lot/certainly, n (%)	Don't know, n (%)
Trust						
I have trust in the health care system.	7 (0.7)	35 (3.4)	102 (9.9)	438 (42.6)	424 (41.3)	21 (2.1)
I have trust in physicians.	2 (0.2)	29 (2.8)	72 (7.0)	412 (40.1)	502 (48.9)	10 (1.0)
I have trust in technology.	4 (0.4)	32 (3.1)	129 (12.6)	519 (50.5)	313 (30.5)	30 (2.9)
Fear						
I believe that AI will lead to unemployment.	122 (11.9)	216 (21.0)	251 (24.4)	169 (16.5)	95 (9.3)	174 (16.9)
I believe that AI will cause unintentional harm to humans.	55 (5.4)	206 (20.1)	303 (29.5)	181 (17.6)	67 (6.5)	215 (20.9)
I believe that AI will lead to loss of control to machines.	86 (8.4)	167 (16.3)	249 (24.2)	241 (23.5)	141 (13.7)	143 (13.9)
I believe that AI will lead to increased data collection and mass surveillance.	22 (2.1)	30 (2.9)	106 (10.3)	309 (30.1)	435 (42.4)	125 (12.2)
Hope						
I believe that AI will lead to more jobs.	119 (11.6)	180 (17.5)	300 (29.2)	164 (16.0)	55 (5.3)	209 (20.4)
I believe that AI will lead to longer lives.	69 (6.7)	114 (11.1)	243 (23.6)	284 (27.7)	92 (9.0)	225 (21.9)
I believe that AI will lead to more quality of life.	82 (8.0)	119 (11.6)	279 (27.2)	265 (25.8)	93 (9.0)	189 (18.4)
I believe that AI will lead to peace and political stability.	225 (21.9)	209 (20.4)	232 (22.6)	66 (6.4)	20 (1.9)	275 (26.8)

[a]AI: artificial intelligence.

Table 3. Respondent characteristics and the importance of attributes.[a]

Attribute (average weight)	Gender	Age	Level of education	Urban/rural background	Chronic disease	Inpatient last year	GP[b] visits last year	Trust scale	Fear scale	Hope scale
Type (0.12268)	—[c]	—	—	—	—	—	—	—	—	• *P*=.002 • r=–.097
Explanation (1.10638)	—	—	—	—	—	—	—	—	—	—
Performance (0.31895)	• *P*=.01 • M[d]=.337 • F[e]=.300	• *P*<.001 • More important with lower age	• *P*<.001 • Lowest level of education=.179 • Highest level of education=.659	• *P*=.002 • Most rural=.271 • Most urban=.354	—	—	—	• *P*=.003 • r=.093	• *P*<.001 • r=.170	• *P*<.001 • r=.243
Responsibility (1.90018)	—	—	—	• *P*=.01 • Most rural=1.940 • Most urban=1.792	—	—	—	—	—	—
Discrimination (0.60190)	• *P*<.001 • M=.542 • F=.682	—	—	—	—	—	—	—	—	• *P*<.001 • r=–.120
Severity of disease (0.06042)	—	—	• *P*=.01 • Lowest level of education=.122 • Highest level of education=–.226	—	—	—	—	• *P*=.002 • r=–.099	—	• *P*<.001 • r=–.168

[a]Numerical data only shown for cells where there is a statistically significant difference.

[b]GP: general practitioner.

[c]Not applicable.

[d]M: male.

[e]F: female.

The results in Table 1 show that the physician having the final responsibility for treatment choice is the most important attribute, with 46.8% of the total weight being allocated to it, followed by explainability of the decision (27.3%) and whether the AI system has been tested for discrimination (14.8%). These 3 attributes accounted for 88.9% of the total weight/importance.

As can be seen in Table 2, the respondents in general trusted health care and technology; did not particularly fear AI, although some did; and in general believed that AI will have positive implications for society.

Table 3 shows that while gender, age, level of education, whether respondents live rurally or in towns, their trust in health and technology, and fear and hope regarding AI did influence the importance they allocated to different attributes, they did not play a significant role in the majority of cases. The data shown for the numerical differences between groups and the correlation coefficients indicate that the statistically significant findings do not reflect large numerical differences or strong correlations.

Discussion

Principal Findings

The results of this study are interesting. First, the study shows that among the respondents, there was a clear order of preference between AI performance and AI explainabillity. Being provided with an explanation was the second-most important factor (27.3%) for the respondents' choice of preferred AI system, while performance carried little weight (6.6%) and was ranked only fourth out of the 6 attributes. However, the study also shows that the population finds a number of different aspects

of AI decisions important for their choice of preferred AI system use. It is not simply a matter of choosing between the performance and explainability of AI. The single-most important factor for their choice is how responsibility for a diagnosis or treatment plan is distributed between physician and AI system. The respondents placed significant emphasis on physicians being responsible for health care decisions (46.8%).

The relatively limited role of AI system performance in the respondents' preferences arguably reflects the chosen levels of the performance attribute. We did not in this study include the possibility that the AI system performs worse or significantly worse than physicians. This would likely have changed the overall weight of performance in the respondents' decisions. Our design was based on what we believe to be the most likely future scenario for the implementation of AI in health care, and this does not include the introduction of AI systems that perform significantly worse than physicians. Implementing such suboptimal systems is likely to be resisted by health care professionals and will in some jurisdictions also be open to legal challenge. Our study specifically shows that in health care implementations with AI systems performing at least as well as physicians, the role of AI system performance in the populations' preferences is limited. They are not particularly interested in getting increased performance if this leads to a loss of other important features of the AI system. On a more speculative note, the choice of a nonnumeric description of the standard of performance ("equally good," "somewhat better," and "significantly better") may be thought to be less informative than providing, for instance, the accuracy of the AI system and physicians as a percentage of correct decisions made, false positives, false negatives, etc. However, providing the information about accuracy in percentage terms may be difficult to understand and may communicate a false sense of precision in our evaluation of how well a system works when implemented in a routine health care setting.

Interpreting the respondents' strong preference for the explainability of AI decisions in health care is difficult. The explainability attribute is stated in terms of the degree of explainability relative to the explanation provided by a physician, and this entails that little can be said about what kind of explanation the respondents want or how they understand explainability. It may be reasonable to assume, however, that this standard would lead the respondents to expect limits to how fine-grained explanations of AI decision making can be made available—just as there are limits to how fine-grained explanations physicians can provide. Thus, providing patients with information about each and every aspect of diagnostics or treatment planning is not the standard of everyday clinical practice for a number of different reasons, including limits to the amount of medical information patients may be able to process and time constraints on the physician-patient encounter. We believe that the relative standard of explainability is the simplest and most meaningful way of introducing different levels of explainability of AI decision making to the respondents.

There are a number of statistically significant findings of relationships between respondent characteristics and the weight given to particular attributes (Table 3). However, when considering the size of the differences in weighting between the highest and the lowest group, it is evident that although these differences are statistically significant, they do not signify a fundamental change in the ranking of the different attributes. The negative finding that there are no significant relationships between the degree of contact with the health service and the weight given to the different attributes is also important since it indicates that there is, broadly speaking, no major differences between patients and nonpatients in their general preferences in relation to AI in health care.

Table 3 contains 2 further key findings. First, that hope in the future benefits of AI certainly is a driver of respondents' views as to the importance of the attributes. The performance of the AI system is of greater importance for the hopeful, whereas testing for discrimination and distinguishing between the use of AI systems for diagnostics or treatment of less severe or more severe diseases is of lesser importance. Second, that the respondents' views concerning the importance of performance are influenced by a number of factors. It is significantly more important for the younger, the educated, the urban, and respondents with fear or hope concerning a future with AI.

Strengths and Limitations of This Study

The response rate of 61.2% is good for a population survey, and the sample efficiency analysis shows that the respondents were similar to the complete sample in relation to the stratification variables. There is therefore reason to believe that the findings reflect the views of the wider Danish population.

The choice situation is hypothetical in 2 ways: (1) The respondents are not in an actual situation demanding a choice between AI systems, and (2) currently, the level of use of AI systems for diagnostic and treatment-planning purposes in Danish health care is not as advanced as the scenarios suggest. The first of these abstractions is a feature of the conjoint choice methodology, and the second is a feature of the current level of penetration of AI in health care. The choice situation is, however, close to a clinical situation of which many respondents will have actual experience. For the purpose of deciding a general policy of transparency of AI health care decision making, we believe it is important to know the wider population's preferences as abstracted from the distress of a real-life choice situation. A general policy should, however, consider the diversity of views and preferences of the general population, including those of patients. Of the 1027 respondents in this study, 375 (36.5%) indicated that they have a chronic disease, 146 (14.2%) had been hospitalized within the past year, and 830 (80.8%) had visited their GP within the past year. Most of the respondents are thus in regular contact with the health care system and are used to being involved in decision making in that context.

Given the rapid developments in AI over the past decade, the general population's familiarity with the potential and actual use of AI systems for diagnostics and treatment planning may be expected to be low. Even those who realize that the voice recognition functionality in their phone or tablet relies on AI processing may not transfer that to the health care context. Asking for views and preferences in relation to hypothetical implementations of AI in hospitals therefore is likely to reflect

more general views and preferences in relation to the implementation of "a new technology" in health care. These views and preferences may or may not change as the familiarity with the technology grows. Potential changes in views and preferences must be monitored and considered in an ongoing adjustment of policies. Such changes do not, however, obviate the need for policy decisions at the current stage of AI development and use in health care.

This entails that it is important that the choice situation be described in a way that does not introduce overt bias, especially in relation to the features of AI that are the attributes of the options in the choice task. There are 2 elements of our choice task that could potentially introduce bias. The first, which we discussed above, is that the performance of AI is compared to the performance of a physician and that choices therefore to some extent depend on the respondents' perception of the typical performance of physicians. The second is our description of the transparency and performance of AI systems. We wrote, "AI systems have shown themselves to be quite good at making diagnoses and suggestions about treatment. However, they also sometimes make errors or differentiate unjustly between patients. In addition, it can be difficult to explain how a diagnosis or a suggestion for treatment has been derived." This is a true description of current and near-future AI systems in health care and is not overtly biased. As all short descriptions, it can be made longer and complete in various ways, but putative proponents or opponents of AI in health care are likely to look for different additions and likely to point to different real-world examples, for instance, in relation to the risk of bias and discrimination. We cannot rule out bias, partly because there is no yardstick for a "neutral description," but we would argue that the risk of bias is low.

This study was conducted in the Danish population and may not be representative of other populations. Further studies are required.

Previous Empirical Research

A number of studies on the broader public's perceptions of automated decision making and AI as such have been conducted in recent years [15-18]. Of particular interest is a recent study on the relative weight of 3 AI features (performance, explainability, and the effort required for implementation/training) for industry experts' choice of a preferred AI decision support system for high-stake maintenance of airplane turbines. The study found that performance is, by far, the most important factor (0.61), whereas explainability (0.20) and effort (0.19) are on a par [19]. Although the study explores the relative weight of AI system features for decisions of a group of experts, it does not report the wider public's perceptions. Moreover, it includes a limited set of features, and it is outside the health care context studied in this article.

Studies of perceptions of AI use in the medical context have taken different approaches. Some have studied the perceptions of AI use among health care professionals [20-23]. We focused here, however, on the public or patient perceptions of AI in the health care context. Overall, the studies on public or patient perceptions report a strong confidence in and acceptability of AI system health care use in the diagnostic context [24-30]. However, most studies also find that the respondents have higher confidence in physician diagnostics and prefer implementations of AI in medical care, with AI playing a decision-supportive role [24,25,27,28]. Several studies indicate that respondents have a strong interest in the performance or accuracy of AI diagnostic systems [28,31]. One study reports a marked difference in respondents' confidence between AI use for diagnostic and treatment decision purposes, with a significant lower confidence in the latter [25]. A study of factors driving the perceived risks of AI clinical device use concludes that perceived uncertainties about performance, concerns about potentially reduced communication with physicians, perceived untrustworthiness of AI, perceived lack of regulatory standards for evaluating the safety and impact of AI, and concerns about liability issues are significant drivers of the perceived risk [31]. Notably, the study also finds that privacy concerns and concerns about social biases and discrimination do not significantly impact the perceived risk of AI clinical device use [31]. Although these previous studies concern perceptions of several features of AI health care use, none of them investigate public perceptions of the importance of AI explainability and perceptions of the relative weight of features of AI decision making in health care. Most of the findings are, however, compatible with our findings.

Policy Implications: The Use of AI in Health Care and Requirements of Transparency

Deciding issues of public policy cannot be done entirely on the basis of individual or population preferences. Various other ethical, legal, professional, and political concerns must be considered. There are, however, at least 2 reasons for considering individual and population preferences. First, the principle of respecting and promoting individual autonomy is usually taken to entail that it counts in favor of an action or a policy if it provides individuals with an opportunity to protect and pursue their interests. Second, the ideal of representative democracy is usually taken to entail that in deciding public policies, decision makers should represent the interests of the people. Designing public policy partly on the basis of population preferences, as mapped in this conjoint analysis survey, is a way of considering individuals' interest at an aggregate level.

This study is of relevance for policies concerning the implementation of AI in health care. The study suggests that AI systems will be found acceptable and can be used for both diagnostic and treatment-planning purposes regardless of the severity of the medical condition if the system has been tested for discrimination, if decisions can be explained to the same extent as physicians' health care decisions, if the physicians are ultimately responsible for the health care decisions, and if the performance of the AI system is at least as good as that of physicians.

The study also has some implications for the question of the transparency of AI decision making in health care. The study clearly shows that a requirement of transparency cannot be dismissed on the grounds that in the eyes of the population, performance is the only significant concern. On the contrary, the population not only takes explainability to be considerably more important than performance, but it also puts significant

emphasis on other aspects of AI decision making. The mere fact that all aspects of the use of AI were considered somewhat important by the respondents could be taken to support the view that the transparency of AI decision making in health care should concern more than the narrow explainability of the AI decision. Transparency should also be a matter of providing patients with information about responsibility and testing for discrimination, performance, and the character of the use of AI.

Taking a more comprehensive approach to transparency fits recent writings on the ethics of transparency of AI use in health care. Thus, it has been argued that transparency in relation to AI-based diagnostics and treatment planning is a matter of providing patients with information that will enable them to effectively contest these decisions [32]. This includes information not only about the key indicators behind an AI-generated diagnosis or treatment plan but also about the performance of AI, bias testing, and the distribution of responsibility between physicians and AI systems.

Conclusion

This paper proceeded from the assumption that AI system performance and AI explainability/transparency are potentially in conflict. Currently, the decision making of the best-performing deep learning models cannot be fully scrutinized or replicated step-by-step. We believe this tension must be resolved in and through appropriate policy making, and we have argued here that an appropriate policy should consider the population's interests in and views concerning AI system features, such as performance and explainability/transparency. The findings of the choice-based conjoint survey reported in this paper are that if an AI diagnostic system does not perform any worse than a physician, then the 3 factors that are most important to the public are, in descending order of importance, (1) that physicians are ultimately responsible for diagnostics and treatment planning, (2) that the AI decision support is explainable, and (3) that the AI system has been tested for discrimination. A policy of AI system use in the health care setting should give priority to AI systems that can meet these requirements and should provide patients with information about the division of labor and responsibility between physicians and the AI system, the key explanatory factors in the AI system decision making, and the bias testing of the AI system. However, this study links the notion of AI explainability to the explainability of physicians' decision making. How AI system explainability can be achieved in a way that makes it relevantly similar to physician decision making is an obvious avenue for further research.

Conflicts of Interest

None declared.

References

1. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health 2019 Oct;1(6):e271-e297. [doi: 10.1016/s2589-7500(19)30123-2]
2. Lipton ZC. The mythos of model interpretability. Commun. ACM 2018 Sep 26;61(10):36-43. [doi: 10.1145/3233231]
3. Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc 2016 Jan 06;3(1):205395171562251. [doi: 10.1177/2053951715622512]
4. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv:1708.08296v1 2017 Aug 28:1-8 [FREE Full text]
5. Adadi A, Berrada M. Explainable AI for healthcare: from black box to interpretable models. In: Bhateja V, Satapathy SC, Satori H, editors. Embedded Systems and Artificial Intelligence. Singapore: Springer; Apr 08, 2020:327-337.
6. Dam H, Tran T, Ghose A. Explainable software analytics. In: Proceedings of the 40th International Conference on Software Engineering New Ideas and Emerging Results - ICSE-NIER ?18 Internet. Gothenburg, Sweden: ACM Press; 2018 May 27 Presented at: ICSE '18: 40th International Conference on Software Engineering; 27 May-3 June 2018; Gothenburg, Sweden p. 53-56. [doi: 10.1145/3183399.3183424]
7. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep 2019 Jan 21;49(1):15-21. [doi: 10.1002/hast.973] [Medline: 30790315]
8. Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a "right to explanation". AIMag 2017 Oct 02;38(3):50-57. [doi: 10.1609/aimag.v38i3.2741]
9. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019 Sep 02;1(9):389-399. [doi: 10.1038/s42256-019-0088-2]
10. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 119. 2016 May 04. URL: http://data.europa.eu/eli/reg/2016/679/oj [accessed 2020-12-01]
11. Hair J, Anderson R, Tatham R, Black W. Multivariate Data Analysis. 3rd ed. New York: Macmillan; 1992:1-730.
12. Chrzan K, Orme B. An overview and comparison of design strategies for choice-based conjoint analysis. In: Proceedings of the Sawtooth Software Conference. 2000 Jan 01 Presented at: Sawtooth Software Conference; March 2000; Hilton Head Island, SC p. 161-177.

13. Microsoft Denmark. Ny undersøgelse: Danskerne ønsker skarpere regler for kunstig intelligens. 2018 Aug 22. URL: https://news.microsoft.com/da-dk/2018/08/22/ny-undersogelse-danskerne-onsker-skarpere-regler-for-kunstig-intelligens/ [accessed 2020-12-01]
14. Sawtooth Software. The CBC/HB System for Hierarchical Bayes Estimation Version 5. Technical Paper Series, Sawtooth Software, Inc 2009 Aug 20:1-106 [FREE Full text]
15. Araujo T, Helberger N, Kruikemeier S, de Vreese CH. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 2020 Jan 01;35(3):611-623. [doi: 10.1007/s00146-019-00931-w]
16. Kelley P, Yang Y, Heldreth C, Moessner C, Sedley A, Kramm A. Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries. arXiv:2001.00081 [cs] 2019 Dec 27:1-12. [doi: 10.1145/3461702.3462605]
17. Zhang B, Dafoe A. U.S. public opinion on the governance of artificial intelligence. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. USA: Association for Computing Machinery; 2020 Feb 07 Presented at: AIES '20: AAAI/ACM Conference on AI, Ethics, and Society; February 7-9, 2020; New York, USA p. 187-193. [doi: 10.1145/3375627.3375827]
18. Cave S, Coughlan K, Dihal K. Scary robots: examining public responses to AI. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. USA: Association for Computing Machinery; 2019 Jan 27 Presented at: AIES '19: AAAI/ACM Conference on AI, Ethics, and Society; January 27 - 28, 2019; Honolulu, USA p. 331-337. [doi: 10.1145/3306618.3314232]
19. Wanner J, Heinrich K, Janiesch C, Zschech P. How much AI do you require? decision factors for adopting AI technology. In: Proceedings of the 41st International Conference on Information Systems. 2020 Dec 13 Presented at: ICIS 2020: Forty-First International Conference on Information Systems; December 13-16, 2020; Hyderabad, India p. 1-17.
20. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):1-8. [doi: 10.2196/17620]
21. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. Front Artif Intell 2020 Oct 21;3:578983-578987 [FREE Full text] [doi: 10.3389/frai.2020.578983] [Medline: 33733219]
22. Liyanage H, Liaw S, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb Med Inform 2019 Aug 25;28(1):41-46 [FREE Full text] [doi: 10.1055/s-0039-1677901] [Medline: 31022751]
23. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14-13 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
24. Yang K, Zeng Z, Peng H, Jiang Y. Attitudes Of Chinese Cancer Patients Toward The Clinical Use Of Artificial Intelligencecancer patients toward the clinical use of artificial intelligence. PPA 2019 Nov 01;13:1867-1875. [doi: 10.2147/ppa.s225952]
25. York T, Jenney H, Jones G. Clinician and computer: a study on patient perceptions of artificial intelligence in skeletal radiography. BMJ Health Care Inform 2020 Nov 13;27(3):1-7 [FREE Full text] [doi: 10.1136/bmjhci-2020-100233] [Medline: 33187956]
26. Zhang Z, Genc Y, Xing A, Wang D, Fan X, Citardi D. Lay individuals' perceptions of artificial intelligence (AI) - empowered healthcare systems. In: Proc Assoc Inf Sci Technol. 2020 Oct 22 Presented at: 83rd Annual Meeting of the Association for Information Science & Technology; October 25-29, 2020; Pittsburgh, USA p. 1-9. [doi: 10.1002/pra2.326]
27. Ongena YP, Yakar D, Haan M, Kwee TC. Artificial intelligence in screening mammography: a population survey of women's preferences. J Am Coll Radiol 2021 Jan;18(1 Pt A):79-86 [FREE Full text] [doi: 10.1016/j.jacr.2020.09.042] [Medline: 33058789]
28. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May 01;156(5):501-512 [FREE Full text] [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
29. Stai B, Heller N, McSweeney S, Rickman J, Blake P, Vasdev R, et al. Public perceptions of artificial intelligence and robotics in medicine. J Endourol 2020 Oct 01;34(10):1041-1048 [FREE Full text] [doi: 10.1089/end.2020.0137] [Medline: 32611217]
30. Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social media. J Med Internet Res 2020 Jul 13;22(7):e16649-e16611 [FREE Full text] [doi: 10.2196/16649] [Medline: 32673231]
31. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020 Jul 22;20(1):170-119 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
32. Ploug T, Holm S. The four dimensions of contestable AI diagnostics: a patient-centric approach to explainable AI. Artif Intell Med 2020 Jul;107:101901-101905 [FREE Full text] [doi: 10.1016/j.artmed.2020.101901] [Medline: 32828448]

Abbreviations

AI: artificial intelligence
GDPR: General Data Protection Regulation

GP: general practitioner

Edited by G Eysenbach; submitted 18.12.20; peer-reviewed by M Wright, R Boumans; comments to author 19.05.21; revised version received 31.05.21; accepted 11.11.21; published 13.12.21.

Please cite as:
Ploug T, Sundby A, Moeslund TB, Holm S
Population Preferences for Performance and Explainability of Artificial Intelligence in Health Care: Choice-Based Conjoint Survey
J Med Internet Res 2021;23(12):e26611
URL: https://www.jmir.org/2021/12/e26611
doi:10.2196/26611
PMID:34898454

Original Paper

Use of Natural Spoken Language With Automated Mapping of Self-reported Food Intake to Food Composition Data for Low-Burden Real-time Dietary Assessment: Method Comparison Study

Salima Taylor[1*], MS; Mandy Korpusik[2*], PhD; Sai Das[1], PhD; Cheryl Gilhooly[1], PhD; Ryan Simpson[3], MS; James Glass[2], PhD; Susan Roberts[1], PhD

[1]Jean Mayer United States Department of Agriculture Human Nutrition Research Center on Aging, Tufts University, Boston, MA, United States

[2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Susan Roberts, PhD
Jean Mayer United States Department of Agriculture Human Nutrition Research Center on Aging
Tufts University
711 Washington Street
Boston, MA, 02111
United States
Phone: 1 617 556 3238
Email: susan.roberts@tufts.edu

Abstract

Background: Self-monitoring food intake is a cornerstone of national recommendations for health, but existing apps for this purpose are burdensome for users and researchers, which limits use.

Objective: We developed and pilot tested a new app (COCO Nutritionist) that combines speech understanding technology with technologies for mapping foods to appropriate food composition codes in national databases, for lower-burden and automated nutritional analysis of self-reported dietary intake.

Methods: COCO was compared with the multiple-pass, interviewer-administered 24-hour recall method for assessment of energy intake. COCO was used for 5 consecutive days, and 24-hour dietary recalls were obtained for two of the days. Participants were 35 women and men with a mean age of 28 (range 20-58) years and mean BMI of 24 (range 17-48) kg/m^2.

Results: There was no significant difference in energy intake between values obtained by COCO and 24-hour recall for days when both methods were used (mean 2092, SD 1044 kcal versus mean 2030, SD 687 kcal, *P*=.70). There were also no significant differences between the methods for percent of energy from protein, carbohydrate, and fat (*P*=.27-.89), and no trend in energy intake obtained with COCO over the entire 5-day study period (*P*=.19).

Conclusions: This first demonstration of a dietary assessment method using natural spoken language to map reported foods to food composition codes demonstrates a promising new approach to automate assessments of dietary intake.

(J Med Internet Res 2021;23(12):e26988) doi:10.2196/26988

KEYWORDS

energy intake; macronutrient intakes; 24-hour recall; machine learning; convolutional neural networks; nutrition; diet; app; natural language processing

Introduction

National recommendations encourage self-monitoring of energy intake to support healthy weight management [1-3]. Mobile phone–based apps are widespread and include features like food imaging and portion guides [4-6]. However, none of the available apps appear to be more accurate or decrease user burden compared with earlier methods [7], and high user burden

combined with limited accuracy are major barriers to routine use. Moreover, high user burden is associated with modifications to usual eating habits that result in atypical energy and nutrient intakes during the measurement period [8].

The use of speech understanding technology for the assessment of food intake is in its infancy but has tremendous potential to reduce the burden of dietary assessment and increase method accuracy, in particular when combined with technologies to map reported foods to their appropriate food composition code for automated nutritional analysis. We have developed an app called COCO (Conversational Calorie Counter) Nutritionist [9], which combines natural spoken language with machine learning to enhance the capture of self-reported dietary intake and map it to food codes in widely used food databases. The primary goal of this study was to conduct a feasibility study evaluating COCO for measurement of energy intake compared to a recommended method [10].

Methods

Overview

This paper describes the first evaluation of COCO for automated self-reported dietary intake. The evaluation was conducted by nutrition research staff at the Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University in Boston, Massachusetts, with data collection and analysis implemented by individuals who had no role in the development of COCO.

COCO Nutritionist

COCO was developed primarily by two members of the research team (JG and MK) in collaboration with the nutrition researchers (SBR, SKD, and CG). A detailed description of the development of the app is provided in Multimedia Appendix 1. Briefly, COCO uses machine learning, specifically deep convolutional neural networks (CNN) [9], to accomplish two tasks: (1) tag each word in a user's natural language description of an eating event as a specific food/drink, with its quantity and brand name and/or description [11]; and (2) map each food item to a matching entity in one of several national food databases [12]. The databases used were the United States Department of Agriculture standard reference and branded foods [13], the University of Minnesota database [14], foods scraped from restaurant websites, and the Nutritionix database [15]. A front-end iOS app [16] makes calls to a back-end Python Flask [17] server that runs the trained CNN models.

Food database measuring units (eg, cup) and amount units (eg, 1, 2) are ranked using string matching. For example, if the user says "cup," it will rank the unit "cup" as the top match. If there are no exact matches, then it defaults to the unit that the user logged for that food item most often previously; if the user never ate that food before, then it defaults to the most popular unit logged by all users. Each user's food logs are stored in a PostgreSQL [18] database on Amazon Web Services [19]. Thus, a user's subsequent logs are customized such that the paired food codes and amounts of user-described foods they have eaten previously are ranked higher or at the top. The input is a spoken or written meal description (eg, "For dinner, I had a bowl of chili with cheddar cheese"). The output is a list of top-15 matching database foods per tagged food item (ranked according to the dot product similarity scores between a vector representing the user's description of the food they ate and a vector representing each food database entry), along with their ranked quantities (ranked according to keyword matching), predicted number amount (eg, 1, if the user logged "a cup"), and corresponding nutrition facts from the selected nutrition database, calculated for each food item. If there are duplicates for foods in different databases, the current system will return both, ranking based on matching the wording in the database versus the user's language.

Participants implicitly confirmed the top-1 food code and food amount proposed by COCO, by not making any changes, or they picked a more suitable food code from the database if the top food code was not the right match, by scrolling through a list of 15 options. If none of the presented options were a good match, participants could delete the food and repeat. A similar process was used for food amounts.

Evaluation of Participants and Informed Consent

Participants and Informed Consent

The research protocol for evaluation of the new COCO app was reviewed and approved by the Tufts Health Sciences Institutional Review Board, and informed consent was obtained from each participant prior to enrollment. In total, 35 participants were recruited using flyers, word of mouth, and emails to a mailing list of individuals requesting to be contacted about nutrition research study opportunities or expressing interest in the app. To be eligible, participants had to be at least 18 years old and generally in good health. Individuals were ineligible if they did not have an iPhone with iOS 10.2 or higher, were unwilling to log their food intake for 5 consecutive days and complete 2 diet recalls, or had a graduate degree in nutrition. Enrollment occurred between June 2018 and June 2019.

Evaluation Protocol

We developed an evaluation protocol for the COCO prototype that was consistent with recommendations for development of new food intake assessment tools [20]. Participants were enrolled in a 5-day study, and completed all components remotely. They were provided with a written overview of the study, including instructions on how to download and use the COCO app, as well as a short video providing tips on use. Participants were asked to use COCO to record the amounts and types of all consumed food and drinks for 5 consecutive days. The recording of intake could be completed using natural spoken language utterances or manual text entry. If a day of recording was missed, they were asked to log an additional day to complete a total of 5 days of entries. A member of the nutrition team (ST) called participants to schedule 2 calls for the 24-hour food recalls (one per day) on day 3 and later. The diet recalls were rescheduled if the previous day's food logging was missed. Each 24-hour dietary recall took approximately an hour to complete. Following completion of both diet recalls, participants were instructed to complete food logging for the remainder of the study.

The diet recalls used the recommended multiple-pass interviewer-administered method [10], and started with an uninterrupted listing by the participant of all foods and beverages consumed the previous day. In addition, the interviewer went through a forgotten foods list, querying the subject on categories of foods that have been documented as frequently forgotten, followed by a final probe listing all the food items consumed, confirming details of the food and serving sizes. A standard food amount booklet [21] was used to assist in portion size estimation of consumed foods. It has eight sections: (1) the forgotten foods list, (2) glasses and mugs, (3) bowls, (4) mounds, (5) circles, (6) grid and thickness blocks, (7) wedges, and (8) shapes and chicken pieces. The 24-hour diet recalls were analyzed using Food Processor (version 10.13.1; ESHA Research Inc).

Demographic questions were also completed by participants, including for age, sex, race, ethnicity, education, weight, and height. A study satisfaction survey on the use of the app was sent to participants to complete at the end of the study, and a US $20 gift card was mailed to their home address.

Analysis

Data from COCO on energy and macronutrient intakes by study day, amounts of information collected via text versus voice data capture, and percentage of food codes proposed by COCO and revised by participants were captured by the Massachusetts Institute of Technology development team without access to the 24-hour dietary recall data, and provided to the Tufts team for statistical comparison of the methods.

Data for the two days when both methods were performed were averaged for each method. Reported mean energy intakes and percent of energy from protein, fat, carbohydrate, and alcohol were compared between the two methods using paired samples *t* tests and Pearson correlations. In addition, for the COCO data obtained over 5 days, the time-distributed effects of energy consumption were evaluated in relation to the sequence of recalls and the day of the week using multiple pairwise comparisons of means with a Bonferroni correction. For comparisons across the number of recalls, we also used multiple comparisons of means, as well as simple linear regression with and without adjusting for weekday versus weekend effects, to examine time series trends in average energy consumption by the numbered recall performed.

All tests were performed for biologically plausible observations, defined a priori as 2-day average energy consumptions <5000 kcals (1 participant was excluded for implausibility). Analyses were performed using STATA (release 15; StataCorp LLC). The significance level was set at $\alpha \leq .05$.

Results

Participants were 25 females and 9 males with a mean age of 28 (range 20-58) years, and mean BMI of 24 (range 17-48) kg/m^2 (Table S1 in Multimedia Appendix 1). When comparing the 2-day averages of COCO and 24-hour recall, we found no significant difference in energy intake (mean 2092, SD 1044 versus mean 2030, SD 687 kcal/d; *P*=.70) or percent energy from protein (mean 16, SD 3 versus mean 17, SD 5; *P*=.54), fat (mean 35, SD 9 versus mean 36, SD 11; *P*=.27), or carbohydrate (mean 50, SD 11 versus mean 50, SD 9; *P*=.89) between the 2-day comparison of COCO and the 24-hour recall method (Table 1).

Table 1. Energy and macronutrient intakes and number of eating events reported with COCO Nutritionist and 24-hour recall (mean of the same two days for each method).[a]

Variables	24-hour recall (N=34), mean (SD)	COCO Nutritionist (N=34), mean (SD)	*P* value
Energy intake, kcal/day	2030 (687)	2092 (1044)	.70
Percent energy from protein	17 (5)	16 (3)	.54
Percent energy from fat	36 (11)	35 (9)	.27
Percent energy from carbohydrate	50 (9)	50 (11)	.89
Number of meals and snacks	4 (1)	4 (1)	.37

[a]Paired *t* tests were used to compare energy and macronutrient intakes across methods.

Mean values were very similar between the methods, as was the number of reported eating events (meals and snacks). On average, there were 4 items in the 24-hour recall that were not reported in COCO (components of composite food items and beverages), and 2 items in COCO that were not reported in the 24-hour recall (snacks, candy, and beverages), reflecting in part the different ways that composite foods (eg, a hamburger, consisting of meat, bread, and other items) were processed by the two methods. There were also significant Pearson associations between 2-day averages of COCO and 24-hour recalls for percent energy from protein (ρ=0.66; *P*<.001), carbohydrate (ρ=0.58; *P*<.01), and fat (ρ=0.38; *P*=.03), with R^2 values of 0.119-0.438 (Figure S1 in Multimedia Appendix 1).

As shown in Figure 1, there was no significant trend over time in reporting of total energy by COCO (*P*=.69), which remained insignificant when controlling for weekend versus weekday effects (*P*=.73). We additionally evaluated how participants entered data, and the accuracy of mapping foods to appropriate food codes. Most foods were logged by typing rather than by speaking (Table 2). Similar ratios of spoken and written descriptions contained mentions of brands or preparation technique, and the percentage was relatively low at ~15%. Spoken meal descriptions were more likely to mention the units of foods (36.1% versus 24.4%). Furthermore, because the editing function of COCO allowed for participant revision of the food type, brand, or description and unit amount (ie, 1, 2, or 3 cups), how often participants revised the COCO-suggested option was

also tracked (8.1% of the time for food name/description/brand and 27.9% of the time for unit amount).

Figure 1. Energy intake reported over 5 consecutive days through COCO Nutritionist (n=34). There was no significant trend in energy intake reporting (*P*=.19).

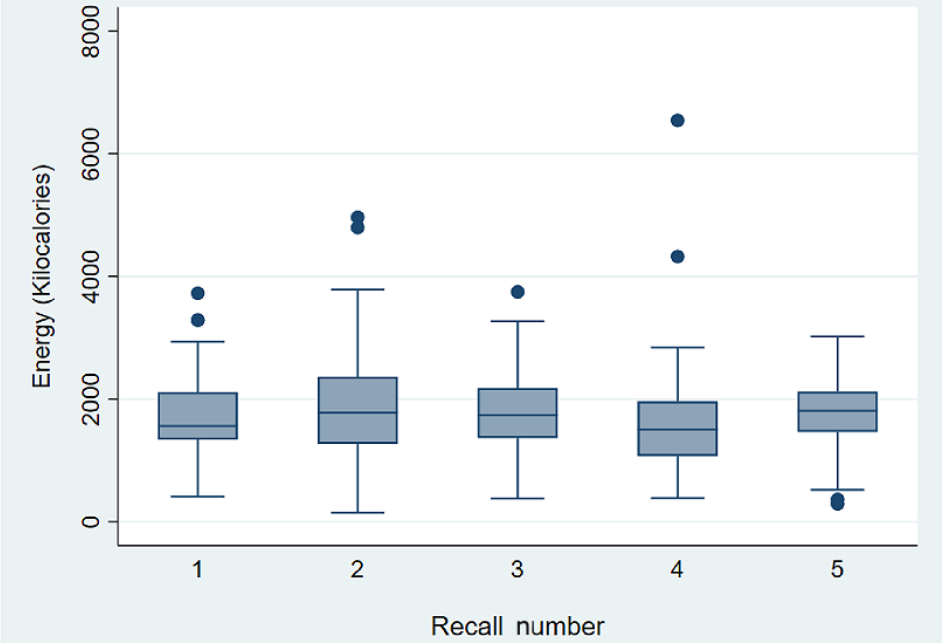


Table 2. Comparison of the food entities logged with spoken versus written natural language in COCO Nutritionist.[a]

Variables	Spoken	Written
Mean number of foods logged per person per day	1.38	7.06
Total of all logged foods, %	28.4	71.6
Mean percentage of foods with specified brands/preparation method	15.6	15.0
Foods with quantities, %	30.7	22.1
Food code identified correctly in default option, %	100	99.9
Food measuring unit (eg, cup) identified correctly in default option, %	83.5	84.1
Food amount unit (eg, 1, 2) identified correctly in default option, %	74.5	84.7

[a]Food code, units, and amounts measured correctly at first try are defined as the user not changing the default.

Discussion

Principal Findings

This study demonstrates that an app using natural spoken language to capture participant self-reports can be combined with automatic mapping of identified foods to information on food composition in national databases to generate estimates of mean self-reported energy and macronutrient intakes—and of the number of meals and snacks eaten—that are not significantly different from those obtained with the gold-standard multiple-pass, interviewer-administered 24-hour recall method. In addition, the lack of any negative trend in reported energy intake by the new COCO app across the 5-day study period was a positive indication that participants were able to complete their reporting without undue burden. To our knowledge, this is the first report of using natural spoken language to collect dietary data in a process that allows for mapping of collected information to food codes to support automated nutrient analyses. Although the data are preliminary, they suggest that this new approach may facilitate accurate, low-cost, and lower-burden methods for dietary tracking for health and weight management.

Comparison With Prior Work

A crucial aspect of this work was demonstration that data collected with natural spoken language can be automatically mapped to a suitable food code in national food databases, something that has not been done previously and allows for automated and instantaneous calculation of dietary intake. This novel feature of COCO will reduce the cost and user burden of dietary assessment, and distinguishes it from previous work using natural spoken language that had a dietician assign food codes to each consumed food [22]. Our demonstration of a relatively low percentage of revisions in the default mapping of collected information to food codes and portion sizes suggests that systems combining voice recognition and writing allow for flexibility in participant use, which may in turn enhance acceptance and sustainable use.

Limitations

Limitations of the study include a relatively small number of features in the COCO app, compared to fully developed apps that include food photography. In addition, the comparator method was self-reported dietary intake by 24-hour recall, which is considered a gold standard for dietary reporting but nevertheless may not reflect absolute dietary intakes. The population sample was relatively small in this first study, and one participant was excluded from data analyses for implausible reporting. Additional work is needed to further refine COCO and to compare collected data with dietary intake of known origin.

Concerning future improvements in COCO focused on making the system as helpful and low-burden as possible, a next step would be to implement a more sophisticated chatbot that can provide personalized food recommendations [23]. For example, the agent could respond if the user asks a question such as "How many calories are in a cup of milk?" or "How many grams of fat have I eaten today?" and recommend foods based on which nutrients the user is missing and their diet preferences. If the user is low in iron, the system could recommend, for example, spinach or steak (both foods high in iron), depending on prior food choices. In addition, taking photos of meals, which is simpler for some users than verbal logging, could provide complementary information to that provided by natural language. Multitask learning research has shown that neural networks trained on multiple input sources and prediction tasks (eg, one side may take an input image, while the other takes an input sentence) have improved performance on both modalities over using each separately [24]. Diet apps use computer vision already; however, to our knowledge, no one has combined language understanding with computer vision.

Summary

Self-monitoring food intake is a cornerstone recommendation in lifestyle interventions for weight management. However, self-monitoring food intake is burdensome with current apps, and frequently inaccurate. This first demonstration of using natural spoken language to map reports of food intake to food composition codes in national food databases and generate automatic assessments of dietary intake demonstrates a promising new approach to substantially reducing user and investigator burden in assessment of dietary intake while supporting the accuracy of reported data.

Acknowledgments

We thank the participants for their participation, and all the Massachusetts Institute of Technology students who contributed to the development of the COCO system, including James Allen, Elizabeth Zhou, Zachary Collins, Rupayan Neogy, Elizabeth Weeks, Liana Reilly, Jiahao Li, Matt McEachern, and Yada Pruksachatkun.

Conflicts of Interest

MK declares that a patent was received for commercialization of the spoken diet tracking algorithm. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary tables, figures, and methods.
[DOCX File , 3177 KB - jmir_v23i12e26988_app1.docx]

References

1. Elfhag K, Rossner S. Who succeeds in maintaining weight loss? A conceptual review of factors associated with weight loss maintenance and weight regain. Obesity Reviews 2005 Feb;6(1):67-85. [doi: 10.1111/j.1467-789x.2005.00170.x]
2. Carter MC, Burley VJ, Cade JE. Weight Loss Associated With Different Patterns of Self-Monitoring Using the Mobile Phone App My Meal Mate. JMIR mHealth uHealth 2017 Mar 02;5(2):e8 [FREE Full text] [doi: 10.2196/mhealth.4520] [Medline: 28153814]
3. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: A randomized pilot study. Obesity (Silver Spring) 2016 Aug 01;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
4. Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. Br J Nutr 2009 Jul;101 Suppl 2:S102-S112. [doi: 10.1017/S0007114509990638] [Medline: 19594959]
5. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Jun;21(8):1495-1502. [doi: 10.1017/S1368980018000393] [Medline: 29534771]
6. Winzer E, Luger M, Schindler K. Using digital photography in a clinical setting: a valid, accurate, and applicable method to assess food intake. Eur J Clin Nutr 2018 Jun 21;72(6):879-887. [doi: 10.1038/s41430-018-0126-x] [Medline: 29563639]
7. Rollo ME, Williams RL, Burrows T, Kirkpatrick SI, Bucher T, Collins CE. What Are They Really Eating? A Review on New Approaches to Dietary Intake Assessment and Validation. Curr Nutr Rep 2016 Oct 15;5(4):307-314. [doi: 10.1007/s13668-016-0182-6]

8. Subar A, Freedman L, Tooze J, Kirkpatrick SI, Boushey C, Neuhouser ML, et al. Addressing Current Criticism Regarding the Value of Self-Report Dietary Data. J Nutr 2015 Dec;145(12):2639-2645 [FREE Full text] [doi: 10.3945/jn.115.219634] [Medline: 26468491]
9. Korpusik M, Glass J. Deep Learning for Database Mapping and Asking Clarification Questions in Dialogue Systems. IEEE/ACM Trans Audio Speech Lang Process 2019 Aug;27(8):1321-1334. [doi: 10.1109/taslp.2019.2918618]
10. Lytle LA, Nicastro HL, Roberts SB, Evans M, Jakicic JM, Laposky AD, et al. Accumulating Data to Optimally Predict Obesity Treatment (ADOPT) Core Measures: Behavioral Domain. Obesity (Silver Spring) 2018 Apr 25;26 Suppl 2:S16-S24 [FREE Full text] [doi: 10.1002/oby.22157] [Medline: 29575782]
11. Korpusik M, Glass J. Spoken Language Understanding for a Nutrition Dialogue System. IEEE/ACM Trans Audio Speech Lang Process 2017 Jul;25(7):1450-1461. [doi: 10.1109/taslp.2017.2694699]
12. Korpusik M, Collins Z, Glass J. Semantic mapping of natural language input to database entries via convolutional neural networks. 2017 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); March 5, 2017; New Orleans p. 2017. [doi: 10.1109/icassp.2017.7953245]
13. FoodData Central. US Department of Agriculture. URL: https://fdc.nal.usda.gov/ [accessed 2021-11-01]
14. Database Licensing. University of Minnesota. URL: http://www.ncc.umn.edu/food-and-nutrient-database/ [accessed 2021-11-01]
15. Nutritionix Database. Nutritionix. URL: https://www.nutritionix.com/database [accessed 2021-11-01]
16. Korpusik M, Taylor S, Das S, Gilhooly C, Roberts S, Glass J. A Food Logging System for iOS with Natural Spoken Language Meal Descriptions (P21-009-19). Current Developments in Nutrition 2019;3(Supplement 1):nzz036. P13-035-19. [doi: 10.1093/cdn/nzz041.P21-009-19]
17. Flask web development, one drop at a time. Pallets. URL: https://flask.palletsprojects.com/en/1.1.x/ [accessed 2021-11-01]
18. PostgreSQL: The World's Most Advanced Open Source Relational Database. PostgreSQL Global Development Group. URL: https://www.postgresql.org/ [accessed 2021-11-01]
19. Amazon Web Services. URL: https://aws.amazon.com/ [accessed 2021-11-01]
20. Eldridge A, Piernas C, Illner A, Gibney M, Gurinović M, de Vries J, et al. Evaluation of New Technology-Based Tools for Dietary Intake Assessment—An ILSI Europe Dietary Intake and Exposure Task Force Evaluation. Nutrients 2018 Dec 28;11(1):55. [doi: 10.3390/nu11010055]
21. Raper N, Perloff B, Ingwersen L, Steinfeldt L, Anand J. An overview of USDA's Dietary Intake Data System. Journal of Food Composition and Analysis 2004 Jun;17(3-4):545-555. [doi: 10.1016/j.jfca.2004.02.013]
22. Rollo ME, Ash S, Lyons-Wall P, Russell A. Trial of a mobile phone method for recording dietary intake in adults with type 2 diabetes: evaluation and implications for future applications. J Telemed Telecare 2011 Aug 15;17(6):318-323. [doi: 10.1258/jtt.2011.100906] [Medline: 21844173]
23. Resnick P, Varian HR. Recommender systems. Commun ACM 1997 Mar;40(3):56-58. [doi: 10.1145/245108.245121]
24. Caruana R. Multitask Learning. Machine Learning 1997;28(1):41-75. [doi: 10.1023/A:1007379606734]

Abbreviations

CNN: convolutional neural network
COCO: Conversational Calorie Counter
USDA: United States Department of Agriculture

Edited by R Kukafka; submitted 27.01.21; peer-reviewed by M Fernandez-Granero, M Majurul Ahsan, C Ochoa-Zezzatti; comments to author 15.04.21; revised version received 02.06.21; accepted 10.11.21; published 06.12.21.

Please cite as:
Taylor S, Korpusik M, Das S, Gilhooly C, Simpson R, Glass J, Roberts S
Use of Natural Spoken Language With Automated Mapping of Self-reported Food Intake to Food Composition Data for Low-Burden Real-time Dietary Assessment: Method Comparison Study
J Med Internet Res 2021;23(12):e26988
URL: https://www.jmir.org/2021/12/e26988
doi:10.2196/26988
PMID:34874885

Original Paper

The Health Care Sector's Experience of Blockchain: A Cross-disciplinary Investigation of Its Real Transformative Potential

Karen Yeung[1], BComm, BCL, LLB, DPhil

Birmingham Law School and School of Computer Science, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Karen Yeung, BComm, BCL, LLB, DPhil
Birmingham Law School and School of Computer Science
University of Birmingham
Edgbaston
Birmingham, B15 2TT
United Kingdom
Phone: 44 0121 414 3344
Email: k.yeung@bham.ac.uk

Abstract

Background: Academic literature highlights blockchain's potential to transform health care, particularly by seamlessly and securely integrating existing *data silos* while enabling patients to exercise automated, fine-grained control over access to their electronic health records. However, no serious scholarly attempt has been made to assess how these technologies have *in fact* been applied to real-world health care contexts.

Objective: The primary aim of this paper is to assess whether blockchain's theoretical potential to deliver transformative benefits to health care is likely to become a reality by undertaking a critical investigation of the health care sector's actual experience of blockchain technologies to date.

Methods: This mixed methods study entailed a series of iterative, in-depth, theoretically oriented, desk-based investigations and 2 focus group investigations. It builds on the findings of a companion research study documenting real-world engagement with blockchain technologies in health care. Data were sourced from academic and gray literature from multiple disciplinary perspectives concerned with the configuration, design, and functionality of blockchain technologies. The analysis proceeded in 3 stages. First, it undertook a qualitative investigation of observed patterns of blockchain for health care engagement to identify the application domains, data-sharing problems, and the challenges encountered to date. Second, it critically compared these experiences with claims about blockchain's potential benefits in health care. Third, it developed a theoretical account of challenges that arise in implementing blockchain in health care contexts, thus providing a firmer foundation for appraising its future prospects in health care.

Results: Health care organizations have actively experimented with blockchain technologies since 2016 and have demonstrated proof of concept for several applications (*use cases*) primarily concerned with administrative data and to facilitate medical research by enabling algorithmic models to be trained on multiple disparately located sets of patient data in a secure, privacy-preserving manner. However, blockchain technology is yet to be implemented at scale in health care, remaining largely in its infancy. These early experiences have demonstrated blockchain's potential to generate meaningful value to health care by facilitating data sharing between organizations in circumstances where computational trust can overcome a lack of social trust that might otherwise prevent valuable cooperation. Although there are genuine prospects of using blockchain to bring about positive transformations in health care, the successful development of blockchain for health care applications faces a number of very significant, multidimensional, and highly complex challenges. Early experience suggests that blockchain is unlikely to rapidly and radically revolutionize health care.

Conclusions: The successful development of blockchain for health care applications faces numerous significant, multidimensional, and complex challenges that will not be easily overcome, suggesting that blockchain technologies are unlikely to revolutionize health care in the near future.

(J Med Internet Res 2021;23(12):e24109) doi:10.2196/24109

KEYWORDS

blockchain; health information management; health information systems; electronic health record; data sharing; health services administration; privacy of patient data; computer security; mobile phone

Introduction

Background

In recent years, distributed ledger technologies commonly referred to as *blockchain* have generated considerable interest and excitement across many industries, including health care, supported by claims of their radically disruptive and transformative potential. Academic literature focusing on blockchain's potential to solve health care problems has proliferated [1-7]. Although scholars have identified many possible applications of blockchain in health care [8], blockchain's perceived value ultimately rests on its potential to create a highly secure, tamper-proof, auditable electronic ledger that can enable responsive, fine-grained, and privacy-respecting access to and sharing of health care data. Accordingly, blockchain is often portrayed as a technological solution that can overcome existing barriers to health information exchange [8], one of the most stubborn challenges that continues to plague contemporary health care [9]. Blockchain advocates claim that these technologies will generate higher quality, more trustworthy, and readily accessible data, which can drive improvements across health care [10]. These improvements would then lead to (1) better quality medical care, including improvements in clinical decision-making and more effective public health management and disease prevention [11]; (2) more efficient, cost-effective, and timely health care administration; and (3) improvements in medical and health care research, resulting from more accurate and secure clinical trial data management and storage [12]. In particular, many believe blockchain will enable fine-grained, patient-controlled access, sharing, and management of electronic health records (EHRs), thereby overcoming existing problems associated with the current *siloed* approach to the storage and management of patient data, which is often regarded as blockchain's *favorite use case* for health care [13,14].

The Blockchain for Health Care Promise: Rhetoric or Reality?

However, expecting blockchain technologies to provide an effective, efficient, patient-centered solution to the multiplicity of problems associated with health care data management and sharing is a tall order. This was one of the expected benefits of EHRs; however, despite their widespread take-up in health care settings, ethically sensitive, lawful, timely, and secure sharing and management of health care data remain a critical but seemingly intractable challenge [9,15,16]. Academic literature concerned with health care applications remains overwhelmingly centered on blockchain's *potential*, focusing on identifying the possible range of benefits that *could* be delivered. However, no serious scholarly attempt has been made to assess the extent to which these technologies have *in fact* been applied to real-world health care contexts and with what consequences, with little attention paid to real-world implementation challenges [17]. In short, academic scholarship remains largely speculative (Multimedia Appendix 1 [1,3,7,18]).

Technical Dimensions of Blockchain Technologies

Overview

To identify what blockchain technologies can and cannot realistically deliver, it is necessary to understand what these technologies are and how they function. Although computer security specialists ascribe a particular technical meaning to *blockchain technologies*, they are understood in a much looser, broader sense for the purposes of this study. In this paper, blockchain technologies refer to a time-stamped database, duplicated across a distributed network of computers (each computer in the network is called a *node*), which is configured such that its technical architecture and operation effectively prevent the rewriting and removal of prior entries. At a basic level, blockchains enable a community of users to record transactions (ie, an interaction between parties) in a shared ledger such that, under the normal operation of the blockchain network, no transaction record can be altered once published. These systems can, as the Bitcoin blockchain demonstrates, enable transactions between strangers without the need for a conventional trusted third-party intermediary, such as a bank or government. Instead, transactions between parties with no pre-existing relationship of trust are enabled by 4 key characteristics of blockchain technologies [19].

Ledger

The technology relies on an append-only ledger to provide a complete transactional history, which, because of its technological design, is almost impossible to amend or alter. This differs from a traditional database in which transactions and entries can be altered or overridden.

Tamper-Proof

Blockchains use cryptographic methods that rely upon mathematical consensus to confirm the consistency and technical authenticity of each transaction's digital record, which is then permanently recorded on the ledger and cannot be altered or deleted. The security and accuracy of the ledger are maintained through the use of cryptographic *keys* and signatures that automatically control who can do what with the ledger. If conflicts between different copies of the database arise (eg, because someone is trying to tamper with the data), the automatic consensus mechanism is designed to ensure that only transactions that are consistent with the earlier, stored version of the database are updated and permanently recorded. Distributed ledger systems that take the blockchain form aggregate transactions into *blocks*, and these are added to a *chain* of existing blocks using a cryptographic signature (hence the name *blockchain*). As the data stored on the ledger can be mathematically attested and cannot be tampered with, it is highly secure.

Shared

A copy of the ledger is replicated and shared across multiple participating nodes, providing transparency across the node participants in the network.

Distributed

As more nodes are added across the network, the ledger becomes increasingly resistant to malicious attacks because it becomes more difficult to interfere with the intended operation of the consensus protocol through which transactions are validated and appended to the ledger. The resilience of the ledger is rooted in its distributed nature. In contrast, conventional databases are based on a centralized, hierarchical design, thus creating a single point of failure.

Blockchain systems can be designed to operate in many ways because of the malleability of the software and protocols through which they are configured. Blockchains may be *permissionless* or *permissioned*, referring to who is entitled to become a network node. Within permissionless blockchain networks, anyone with a computer that has sufficient computing capacity can download the software and participate in the consensus process, storing and updating the shared ledger (referred to as *write* access) without needing permission to do so. As write access is *open* to all for permissionless blockchains, anyone can also *read* or otherwise inspect the ledger. In contrast, permissioned blockchains restrict write access: only those granted advance permission by the relevant authority within the network can participate in the consensus process through which the shared ledger is updated and stored. For permissioned blockchains, the ability of others to *read* and inspect the ledger may be *open* to anyone or can be restricted or *closed* so that only those with authorized access may read its contents. Permissioned blockchains can provide similar functionality to permissionless blockchains but are considerably less computationally demanding. As participating nodes must first be identified and authorized before joining a permissioned network, this provides an assurance of trust across the network, backed by the threat that write access can be withdrawn from misbehaving nodes. As a result, consensus models applied within permissioned blockchains need not provide the same high levels of mathematical and computational security. Therefore, they offer faster performance and are computationally less expensive [19].

The elasticity and malleability of blockchain systems, including the variety of different forms and functionalities that they can offer, has resulted in considerable variation in the way the term *blockchain* is used, particularly in discussions about blockchain for health care. The very broad definition of blockchain adopted here reflects the health care sector use of the term, encompassing any sociotechnical system that uses a distributed, append-only database that relies on cryptographic methods to verify and validate transactions before they are added to the ledger, including both permissioned and permissionless systems, which include distributed ledger technologies that do not aggregate and store data into linked blocks [13].

Aims, Objectives, and Research Questions

The aim of this study is to critically investigate the health care sector's engagement with blockchain technologies to date, analyzing those experiences from a theoretical, cross-disciplinary perspective to ground a more realistic appraisal of the potential benefits, difficulties, and implications of implementing blockchain in health care settings. This investigation was animated by the following research questions:

1. To what extent and in what ways have blockchain technologies been taken up in the health care sector?
2. What are the primary real-world opportunities and challenges that blockchain technology generates in health care contexts?
3. What are the prospects of blockchain take-up and implementation in health care settings to solve real-world problems in the short to medium term?

Taken together, these questions can be amalgamated into a single overarching research question: what does health care's actual experience of blockchain technologies to date reveal about the technologies' *real* revolutionary potential [20]?

This paper proceeds by outlining the study's design, data sources, analytic approach, and theoretical foundations. It then sets out its findings by first providing a qualitative overview and explanation of the extent to which blockchain technologies have been taken up in the health care domain. Second, it offers a critical examination of a series of multidimensional challenges, organized thematically, that constitute (or are likely to constitute) very substantial hurdles that must be overcome if blockchain is to be widely taken up in health care settings. Finally, it discusses these challenges in light of the most promising opportunities for generating real value in health care settings and suggests that the use of blockchain to provide fine-grained patient control over medical records is unlikely to live up to its perceived promise.

Methods

Study Design and Data Sources

The mixed methods study reported in this paper forms the major component of a Wellcome Trust–funded cross-disciplinary project aimed at understanding the challenges, risks, opportunities, and experiences of the health care sector in using blockchain technologies. It took the form of a theoretical investigation, building upon the initial research study by Motsi-Omoijiade and Kharlamov [21] that sought to identify and map evidence of real-world engagement with blockchain technologies in the health care sector, classifying them according to their context of application (as patient care support, administration, or research data management), level of maturity, and the primary geographic region from which the application appeared to originate. The integrated catalog of health care blockchain applications and use cases created in that study provided a snapshot of the state of health care blockchain take-up on November 30, 2019. It identified 128 *applications*, referring to real-world instances in which blockchain technology has been developed into a commercially viable health-related service created largely from a web-based review of English

language websites, gray literature, and academic papers (Multimedia Appendix 2 [21-28]). This study builds on the results of that earlier mapping and classification exercise by drawing on 2 focus group investigations and a series of iterative, in-depth, theoretically oriented desk-based investigations.

Focus Group Workshops

The 2 focus group discussions were organized in which a mix of both academic researchers and blockchain-for-health care sector experts participated. The first workshop, hosted at the outset of the project, was designed to generate insights that could inform and guide the direction of the study's inquiries and help understand the current state of health care sector engagement and experience with blockchain technologies for specific health care purposes. The second workshop occurred toward the end of the study. Its aim was to share the provisional study findings with participants to elicit their critical feedback and provoke discussion. The focus group participants were identified and invited with the aim of bringing wide-ranging, relevant expertise to the discussion, given the project's aim of adopting a multidisciplinary perspective to address the project's research questions. This approach to recruitment reflects a belief that technical, legal, ethical, management, clinical care, and health service industry knowledge and experience would be needed to successfully apply blockchain to real-world settings.

The participant invitation list was compiled from the principal investigator's existing knowledge of academic experts with existing or related interests in blockchain from several disciplinary backgrounds, combined with a light touch review of recent academic and digital health care industry publications to identify key health care sector experts with direct and ongoing experience in seeking to apply blockchain technologies to health care. This initial list was supplemented by the use of snowball sampling techniques (asking each academic and industry invitee to nominate 1 or 2 colleagues) to identify others with relevant expertise. The specific composition of the workshops differed in light of their different objectives and participant availability: the initial workshop comprised more academics than industry experts, whereas the final workshop was attended predominantly by industry experts relative to academics (Multimedia Appendix 3).

Desk-Based Review and Theoretical Investigation

The project's theoretical, desk-based investigations primarily drew on 2 data sources. First, the investigations drew from a wide variety of academic literature from multiple disciplinary perspectives, including computer science, computer security, digital health, organizational management, operations research, medical law, data protection law, medicine, medical sociology, and innovation studies. Given that relatively little information about health-related blockchain projects is publicized via academic journals [29], and existing scholarly literature on blockchain in health care settings has not hitherto been concerned with real-world engagement and experience, relevant academic sources were identified by deductive critical reflection. This entailed developing and refining a series of thematically focused investigations aimed at acquiring a deeper, more fine-grained, cross-disciplinary understanding of the major issues and challenges that would invariably confront those endeavoring to apply blockchain to health care settings. These included matters of data security and integrity, health information technology (IT) implementation, user resistance, internal and external governance, interoperability, organizational management, and data privacy. Second, gray literature (including observations and reflections from those with direct experience of the health sector's engagement with blockchain to date) provided a vitally important source of insight, given that the published scholarly literature represents only a small snapshot of global blockchain activity. This literature included reflections collected from various informal sources, such as guidance and opinion pieces published in various formats, including essay collections [30], blog posts, news articles, newsletters by industry bodies (particularly those published by the Healthcare Information Management and Systems Society [HIMSS]) and key industry experts from organizations offering blockchain for health care consultancy services such as Consensys Health and Hashed Health.

Analytical Approach

This study proceeded from the premise that blockchain's reliance upon cryptographic techniques and networked, distributed computing facilities to provide a highly secure, shared but tamper-proof database means that its foundational value lies in its capacity to facilitate *secure data sharing and collaboration*, particularly between parties in circumstances where a lack of social trust would otherwise inhibit cooperation. So understood, the development, implementation, and scale up of blockchain technologies into health care contexts are likely to pose distinctive and novel challenges because of both its technological design and the health care contexts in which attempts are being made to implement it.

The analysis proceeded in 3 stages. First, it began by seeking to acquire a deeper, more qualitatively oriented understanding of the observed patterns of blockchain for health care engagement mapped from the Motsi-Omoijiade and Kharlamov [21] study. These first-stage inquiries drew heavily on insights and observations contained in industry publications [29], government reports, mainstream media, and initial focus group discussions. The resulting analysis produced a clearer, richer, and more nuanced understanding of (1) how blockchain technologies were intended to provide specific functional capacities across a variety of real-world health care contexts; (2) the range of applications, context-domains, and participants who were expected to engage with and those expected to benefit from these applications; (3) how blockchain was expected to address specific data-sharing problems encountered in health care settings; and (4) the wide-ranging and multidimensional difficulties or *challenges* that have been encountered in the course of these engagements to date or that can be expected to arise if those engagements continue to expand and deepen, particularly if they extend into clinical settings.

The second-stage analysis entailed a critical comparison of the benefits, difficulties, and drawbacks identified in stage 1 against the claims appearing in the academic literature about blockchain's potential benefits and shortcomings (both generally and those specific to health care). This analysis sought to evaluate these claims, which were focused almost exclusively

on the benefits of blockchain, with little mention or consideration of shortcomings, against early-stage experiences with blockchain implementation in real-world health care settings. The resulting mismatch between academic claims and actual experiences of the technology prompted further scrutinization and critical analysis of the implementation problems identified from the first-stage analysis, grouping them together thematically (rather than primarily in relation to specific functional applications or particular kinds of health care data). This grouping was based on recurring difficulties and challenges that have been encountered (or are likely to arise in relation to specific health care applications, particularly those that directly affect clinical practice) in seeking to design, configure, and implement blockchain technologies in health care settings. The encountered challenges were typically multidimensional in character, entailing dynamic interacting and often complex technical, social, regulatory, and organizational difficulties that would need to be addressed successfully before the technology is likely to produce significant value for health care.

The final stage of analysis sought to develop a more theoretically grounded account of the observations generated at the second stage analysis, drawing upon different fields of scholarship, including academic discussions focused on blockchain's *revolutionary* potential; the nature, dimensions, and challenges of innovation more generally; and of health IT implementation in particular (including studies of EHR implementation). The aim was to produce an integrated, theoretically informed, cross-disciplinary understanding of the experience of the health care sector's engagement with blockchain technologies to date, providing a firmer foundation for appraising its prospects by reference to the underlying nature and form of coordination that blockchain technologies currently enable.

Theoretical Frame

The theoretical framework for this study draws primarily on 3 different but complementary perspectives that collectively contribute to addressing the project's overarching research question. First, it draws upon the growing cross-disciplinary literature that has typically emphasized blockchain's *revolutionary* potential to enable novel forms of social cooperation between strangers without the need for a trusted third-party intermediary. Blockchain's potential to enable *trustless* cooperation is ultimately attributable to its underlying technological architecture and distributed operation to produce a shared database that is effectively tamper-proof, enabling the replacement of social trust with computational trust in specific contexts. This has prompted some scholars to describe blockchain as a *truth machine* [31]. More recently, however, a more skeptical strand of literature has started to emerge, drawing attention to a wide range of reasons why the technology may fail to live up to its promise [32]. As these inquiries progressed, it became necessary to look to a wider variety of literature to make sense of these real-world implementation challenges. This led to the investigation of a second body of academic scholarship located within the field of *innovation studies*, a different multidisciplinary strand of work that includes a wide range of industry studies that demonstrate that technologies do not arise in the fully developed form [33-35]. Instead, they find that a period of considerable confusion usually follows the emergence of a new technological invention, with little agreement about what its major subsystems should be or how it can best be put together as a product or service, resulting in significant experimentation [36,37].

Health care's current and ongoing engagements with blockchain can be located within this early experimental phase, situated at the entrance to the so-called *valley of death*. This refers to a *gap* between the development of new scientific knowledge and the establishment of proof of concept but before full development and commercialization [38,39]. Industry estimates suggest that 4 in 5 new inventions are never commercialized because of their failure to overcome numerous *barriers to innovation* encountered during this crucial development phase [40,41]. Innovation studies scholars have identified multiple *valleys of death*, which they have explored from various angles. Some understand the valley primarily as a financing and funding problem that arises after the establishment of a proof of concept. At that moment, significant investment is required to scale up; however, there is considerable technical and commercial uncertainty about its likely success, which thus discourages further investment [34]. Occasionally, this leads to various policy prescriptions, such as government funding initiatives and *translational* centers comprising academic and commercial stakeholders, to help bridge the gap [42]. Others focus on the difficult transition of university research to a product or service brought to market by a commercial firm, particularly in the case of drug development [43].

For the purposes of examining the prospects of blockchain in health care, the most salient strand of this literature is concerned with identifying the *barriers to innovation* that create and contribute to the valley of death. Scholars have classified these barriers as either *internal* to the organization (including factors such as a restrictive mindset, lack of competence, insufficient resources, and an unsupportive organizational structure) or *external* and thus largely beyond the firm's direct control (including factors such as a paucity of external finance; resistance or lack of support from specific actors, such as customer resistance or an unsupportive government; and various factors which result in a restrictive macroenvironment, such as an undeveloped network and ecosystem, technological volatility that narrows the window of opportunity during which an innovation can be introduced, inappropriate infrastructure, and a restrictive local culture) [44]. However, these largely generic, rather abstract, accounts of various obstacles that must be overcome for technological innovations to largely succeed treat the underlying technology as a *black box* rather than engaging with technological design and specific development challenges in seeking to configure and embed the technology into sector-specific social and organizational systems, processes, and practices [45,46].

Rather than treating blockchain in health care as a closed *black-box*, these theoretical investigations drew upon a third set of analytical lenses situated within academic investigations of health care IT system implementations, particularly studies undertaken from a *social practice* perspective [45,47-49]. These studies highlight the critical role of context, culture, and the values that are brought to bear by medical professionals in clinical environments, identifying clinical consultation as a

complex social encounter that occurs within a heavily institutionalized environment [50]. In particular, scholars from these traditions emphasize the role of values in clinicians' understanding of what constitutes and how they seek to practice *excellent care*. Moreover, they stress the nature of clinical knowledge as *tacit, context bound, and ephemeral rather than codifiable, transferable, and enduring* such that the implementation of health IT into clinical settings often departs substantially from those envisaged by system developers and designers [50]. Accordingly, insights of this kind cast significant doubt on the capacity to hard code norms into technological systems in ways that can be seamlessly integrated into clinical environments, especially when these norms concern the sharing of health data, which is typically and legally understood as highly sensitive and worthy of special protection.

Results

Overview

The results of this study are presented in 3 parts, which correspond with the study's analytical approach outlined in previous sections. First, it briefly describes the health care sector's engagement with blockchain technologies to date in the United States, United Kingdom, and at the transnational level. Second, it critically reflects upon a series of 6 challenges that must be overcome and 3 normative trade-offs that must be satisfactorily resolved if blockchain technologies are to cross the *valley of death*, transitioning successfully from *invention* to *innovation* in health care settings. Third, it critically discusses key findings by referring to a selection of wider academic literature concerned with organizational cooperation and radical versus incremental technological innovation and reflects upon the promises of patient-controlled blockchain-enabled EHRs before offering some concluding reflections.

Interrogating Blockchain in Real-world Health Care Contexts

The Motsi-Omoijiade and Kharlamov study [21] revealed that, apart from the Estonian government's eHealth initiative [22], active engagement with blockchain technology in health care began around 2016, primarily in the United States and, to a lesser extent, in Europe, with several recent initiatives occurring at the transnational level. As of November 30, 2019, their study had identified 128 health care blockchain applications (in which a blockchain *application* is defined as providing a specific functional health care purpose so that a single blockchain system might provide multiple applications) from an English-language website review, over half of which were geographically located in the United States (65 applications), with the United Kingdom as the next most popular site of activity (12 applications) [21]. However, less than half of these had a commercially available blockchain product or service, with most still at the experimental or development stage. Of these applications, a substantial minority (50/128, 39.1%) focused on applications to enable user-controlled access to some kind of personal health data, with a significant proportion concerned primarily with using blockchain to facilitate health care administration (34/128, 26.6%) and support for patient care or health management (primarily in facilitating remote care consultations between patient and clinician located elsewhere) or in the form of *wellness* applications to encourage healthy behaviors (31/128, 24.2%) and for medical research data management (13/128, 10.2%; Multimedia Appendix 2).

A closer examination of the technical and contextual dimensions of these applications and their primary sources of funding revealed that, first, the overwhelming majority of blockchain for health care initiatives had used permissioned or hybrid systems, in which the tasks of sharing and updating the ledger are restricted to those authorized to do so. Second, in both the United States and the United Kingdom, private sector investment and funding have been driving these initiatives, although various government programs support blockchain initiatives, including funding support [51,52]. Although some early initiatives sought to generate funding via an *initial coin offering*, including start-up firms claiming specialist blockchain expertise, these have failed to generate positive returns for investors (and a significant number of initial coin offerings generally were merely vehicles for fraud) [23,53]. More recently, investment funding has been provided by venture capitalists, evidencing their confidence in the technology's potential value to the health care sector. The commercial and entrepreneurial expertise that they bring to the sector might help the fledgling industry navigate the fraught and uncertain early development phase [54,55]. Third, the current and ongoing focus of interest in blockchain for health care, at least in the US context, has been predominantly at the intraorganizational and business-to-business (B2B) level rather than being primarily concerned with clinical care provision. Much of this activity has primarily concerned blockchain systems for managing *nonclinical* data, as health care providers explore the capacity to harness blockchain technologies to enhance administrative and operational efficiency. Fourth, by comparison, the UK health sector's experience and engagement with blockchain technology has been far more muted, with only a handful of initiatives, all at the pilot or early-stage implementation phase. These initiatives are primarily for patient care management, with very limited reliance on blockchain functionality. Finally, several promising transnational blockchain initiatives to facilitate medical and health research have been launched in recent years. The following discussion briefly outlines the US and UK experiences (as the 2 most active sites of blockchain engagement in health care) and more recent transnational blockchain initiatives.

The United States

In the United States, ongoing blockchain activity in health care is coalescing at the intraorganizational and B2B level. These applications are intended to facilitate data sharing between business units within the same umbrella health care organization or within recently established consortia of health care organizations seeking to cooperate in the sharing and management of specific forms of health care data stored via a blockchain-enabled ledger [55]. The primary drivers supporting blockchain development to date appear to have focused on its potential to reduce costs and enhance efficiency in health care administration and operations. These applications are broadly concerned with either health care operations management, primarily in the management of supply chains, or administrative data pooling.

Turning first to health care operations management, considerable blockchain activity has occurred either to streamline and automate internal operational processes or to improve the efficiency and verifiability of supply-chain management [29,56]. Examples of the former include the Coalesce Health Alliance consortium's development of blockchain to improve the accuracy and efficiency of patient health care claims management across different entities within the Blue Cross Blue Shield ecosystem [57]. Similarly, the US Department of Health's Office of National Security has built a blockchain providing organizational units with linked, real-time access to a standard set of spend data to improve the department's procurement process, reportedly generating major efficiency gains [58]. In health care supply chain management, significant blockchain activity has occurred to address the potentially serious risks of a compromised health care supply chain [56]. The Food and Drug Administration has actively encouraged these applications by launching a pilot project (announced in February 2019) to encourage drug supply chain stakeholders to develop digital systems to comply with the Drug Supply Chain Security Act 2013, which mandates the creation of an electronic, interoperable system that can trace and identify distributed prescription drugs across the United States. It attracted 26 participants (including several pharmaceutical supply chain stakeholders [59,60]), generating a variety of projects. These include the MediLedger Project, which sought to build a blockchain network around a tamper-proof ledger of pharmaceutical supply-chain transactions to inform responses to product ID verification requests communicated via a permissioned messaging network [61]. The Pharmaceutical Utility Network has taken a different approach, pursuing a *platform first* strategy [61,62] rather than focusing on a specific *use case*. It is developing an open-source blockchain platform that integrates regulatory requirements (such as the Drug Supply Chain Security Act) to enable those involved in the pharmaceutical supply chain (including pharmaceutical manufacturers, distributors, dispensers, and software vendors) to use the platform's services to demonstrate compliance with regulatory requirements, seeking to use blockchain as a form of *RegTech*. [63,64] Other supply-chain management blockchains in the health care sector focus on single use cases, often with some RegTech functionality, such as the Clinical Supply Blockchain Working Group's platform [65], which is now seeking to apply its blockchain-based inventory and event-tracking system to a real clinical trial [66].

Blockchain applications are also being used to facilitate administrative data pooling at the intra- and interorganizational levels. Health care consortia in the United States began emerging in 2018, seeking to use blockchain either to create a single ledger shared between member organizations to serve as a single source of *truth* or to enable network members to create a shared pool of records to which access is controlled and recorded via the blockchain. For example, the Synaptic Health Alliance [67] aims to use a blockchain network to address costs, delays, and inefficiencies in provider directory data (comprising demographic information about physicians and other providers), which insurers are required by federal law to maintain and keep up to date. Each insurer has typically maintained its own directory; however, these often contain inaccurate data, resulting in delayed claims and payment processing. Synaptic's blockchain pilot seeks to create a common source of *truth* for provider directory data created by pooling and comparing each member's directory data to create a complete, accurate shared directory.

Similarly, the Professionals Credentials Exchange (ProCredEx) is a consortium of health care organizations seeking to enhance the efficiency of the physician credentialing process [57]. In the United States, when a new physician is hired or begins working for a new payer's network, the employing health care organization must gather a wide variety of certificates and credentials, a process that takes up to 6 months (sometimes longer), with physicians often required to supply evidence of their professional credentials to a dozen different organizations or more, which must be repeated for all physicians every 2 years by each organization [29]. The ProCredEx blockchain network enables organizations to contribute their credentialing records (essentially, a digital asset) to a shared pool, effectively creating a members-only digital asset exchange so that members can access verified credentials and actively contribute credentials information that other members can acquire. The shared ledger tracks asset ownership, exchange, and use and forms a basis for asset reputation. Buyers of these information assets finance the platform, whereas curators of the assets receive payment for their contributions, as do partners who help facilitate the use of the blockchain network [68,69].

The United Kingdom

In comparison with its US counterpart, interest in blockchain in UK health care has been considerably more limited [70]. The UK blockchain use cases have been directed at patient care management largely outside the clinic, in which secure and auditable data sharing remains the central functionality provided by blockchain technologies. Unlike the United States, UK health care is delivered primarily via a publicly funded National Health Service (NHS), in which NHS hospital treatment and primary care are free at the point of use to UK residents. Decisions about the commissioning of services, including the use of new technologies, are taken primarily at the local level by entities known as clinical commissioning groups or exceptionally at the national level by NHS England, which is informed by guidance issued by the UK's National Institute for Health and Care Excellence [71]. The NHS Long Term Plan states that the NHS will seek to "make better use of data and digital technology" [72] as part of the NHS digital transformation agenda; however, it makes no explicit reference to blockchain technologies. However, the NHS England's Director of Digital Development has been reported as having a general interest in considering blockchain as one of a variety of technologies that the NHS will explore [73].

Nevertheless, NHS England has commissioned Dovetail Labs to develop a novel digital application to enable different members of a multidisciplinary team working with type 2 diabetes patients to share health information. The application operates via a distributed ledger that logs patient consent to share digital data along with user authentication, data access, and transfer records [74,75]. In addition, at least two local clinical commissioning groups have commissioned small-scale

digital applications that are described as using blockchain technology. Hence, a partnership between Guardtime and Instant Access Medical and Healthcare Gateway has produced a comprehensive blockchain-supported personal care record platform (called *MyPCR*). The platform is intended to assist patients in managing chronic long-term health conditions by issuing digital alerts to help them adhere to their personalized treatment plan or personal care pathway via a smartphone app, granting patients complete access to their personal care pathway while generating an automated tamper-proof record of health data provenance and integrity that is compliant with data protection legislation [76]. Similarly, MedicalChain has launched a blockchain pilot enabling patients to create a free wallet to manage access to their digital health records, enable general practitioner video consultations, and enable them to pay for services using cryptocurrency (with users incentivized to pay for telemedicine services with Medicalchain's MedTokens) [77,78].

Transnational Blockchain Systems

At the transnational level, a rather different kind of blockchain application has emerged, which enables network members to run secure but privacy-preserving computational analyses on the data of other members to improve their algorithmic models for research purposes while the data never leaves the host organization [79]. For example, members of the Machine Learning Ledger Orchestration for Drug Discovery blockchain network, comprising 10 large pharmaceutical companies, 5 technical partners, and 2 universities, use blockchain and federated learning to train an algorithm to identify a rare mutation in cancer, enabling pharmaceutical companies to collaborate on therapeutic drug discovery. Each pharmaceutical company can, via the network, run the machine learning algorithms of their academic partners on each other's data sets. The data remain at their local site so that proprietary information from one data set cannot be leaked to another. Nonsensitive algorithmic models are exchanged between the members, which are then collectively consolidated to improve the predictive performance of the algorithmic models by leveraging all the data across the federation. A permissioned blockchain tracks and logs all operations taking place on the federated data using a software framework (Substra) for enabling the execution of distributed machine learning tasks in a secure way. The value of collaboration is particularly evident in attempts to develop drug treatments for rare diseases: each individual organization will have too few patients with a given rare disease to undertake statistically significant analysis; however, by pooling access to their data to enable federated learning, collaboration without data sharing becomes possible. Similarly, the AI Centre for Value-Based Healthcare consortium has created a federated learning network with patient data from 4 hospitals and 3 universities, allowing research partners to train algorithms on the federated data set. Collaborations of this kind, which use blockchain networks to facilitate privacy-preserving, federated data access, offer considerable promise to advance medical research [29]. Health care blockchain industry expert Robert Miller [80] observes the following:

> *Health data is inherently sensitive, and thus demanding of privacy. Yet at the same time health data is inherently statistical, and thus holds within it insights that could improve health for all. The promise of federated learning is to unlock these insights without compromising on privacy. Moreover, federated learning will enable us to assemble more data than ever in federated networks, leading to better algorithms and ultimately better outcomes.*

Challenges for Blockchain Take-up and Implementation in Health Care Settings

Overview

We have seen how the health care sector is beginning to develop viable blockchain applications for specific health care purposes and contexts to establish proof of concept, which are yet to be widely implemented at scale [58]. This invariably invites questions about the prospects of blockchain's sector-wide take-up, implementation, and diffusion, particularly given the limited success of health IT adoption more generally [81]. To begin addressing these questions, a set of 6 interrelated *implementation challenges* associated with seeking to implement and integrate blockchain into real-world health care contexts were identified using the methodological approach described above. Of these, 3 challenges were already being encountered in seeking to use blockchain technologies to facilitate the sharing and collaboration of health care data for administrative and research purposes, namely (1) organizational commitment, (2) interoperability and standardization, and (3) internal governance. A total of 3 additional emerging challenges are likely to be particularly acute if, in the future, sustained attempts are made to introduce blockchain technologies into clinical settings for which only modest experimentation has occurred to date, namely (4) data security and integrity, (5) quality and safety, and (6) truth and immutability (Textbox 1).

In addition to these implementation challenges, 3 further *normative tensions* or value conflicts invariably arise in configuring blockchain technologies to meet specific functional requirements in health care contexts, which were also identified (Textbox 2). Accordingly, acceptable compromises between competing objectives and values must be identified in an attempt to apply blockchain to real-world settings that can be designed into the software and system architecture and implemented into specific organizational contexts [82].

Textbox 1. Blockchain in health care implementation challenges.

Implementation challenges

- Organizational commitment
- Interoperability
- Internal governance and standardization
- Data security and integrity
- Quality and safety
- Truth and immutability

Textbox 2. Normative tensions requiring resolution in applying blockchain to health care.

Normative tensions

- High performance and scalability while providing adequate security
- Providing transparency and accountability while ensuring due respect for privacy and confidentiality
- Establishing computational trust while securing adequate social trust between network participants

Common to each of these challenges and normative tensions is their multifaceted, dynamic, context-sensitive, and complex character, comprising both technical and nontechnical dimensions. Moreover, addressing these challenges typically requires internal organizational adaptations *and* changes external to the organization because of the inherent character of blockchain as a *process* innovation [83], particularly when used to facilitate trustworthy data sharing beyond organizational boundaries. Each of these challenges is outlined below, beginning with 3 near-term challenges, followed by a brief discussion of 3 further challenges that are likely to be particularly acute if blockchain applications are applied in clinical contexts.

Organizational Commitment

The decision by any organization to adopt a new IT system is not made lightly. Making a commitment to experiment with a technology to establish proof of concept as a bounded, time-limited project is likely to be considerably easier for an organization than committing to implementing a novel technology on an ongoing basis, particularly for health care organizations [84]. Difficulties associated with eliciting the funding and commitment necessary to move a technological invention beyond the proof-of-concept stage are one of the characteristic features of the so-called valley of death described above. For organizations, this requires a compelling business case demonstrating that such a move will generate significant, sustainable net gains (such as increased revenues, reduced costs, or better-quality patient care) to justify the substantial investment and risks associated with doing so [85,86]. Given that the nature and magnitude of benefits of blockchain for health care remain largely unproven, and experience has shown that implementing health IT programs involves a great deal of time, money, and effort, making such a case will be difficult [9,81,85,87]. When proposed to support administrative functions, such as supply-chain management, the cost of blockchain adoption and implementation may be difficult for an organization to justify, as the expected benefits take the form of enhanced compliance and, hopefully, lower costs and reduced risk, all of which are difficult to quantify. For example, consider difficulties in quantifying the value gained from using blockchain-enabled pharmaceutical supply chains to reduce the risks associated with the production of fake medicines whose scope and prevalence are not well-known, and alternative cheaper technologies which may be more attractive [29]. Organizations willing to attempt blockchain implementations will also need to join a suitable network of participants willing to align and collaborate over a common interest in a particular kind of health care data via a blockchain network and avoid generating significant user resistance from staff. Neither task is likely to be easy [30,32,88-90].

Nonetheless, the proliferation of health care consortia in the United States, throughout 2019 and early 2020, seeking to develop blockchain-based collaborations over administrative data suggests that a variety of health care organizations increasingly recognize the potential value of blockchain-based data sharing for administrative and research purposes. The use of blockchain-enabled data sharing to support highly labor-and time-intensive administrative processes (such as provider credentialing) could significantly enhance administrative efficiency and reduce workloads. As one leading commentator has observed, almost every insurance company and pharmaceutical company has announced participation in a cooperative data-sharing health care consortium, including health systems, IT companies, and nontraditional health care companies such as PNC Bank and Walmart [91]. However, blockchain applications that have gained the most organizational buy-in have either avoided sharing of patient data, focusing instead on collaboration over administrative data, or have involved research collaborations that share and manage patient data in a privacy-preserving manner.

This points to a further hurdle that is likely to arise in eliciting organizational commitment: the need to ensure that blockchain applications comply with applicable legal and regulatory requirements, which are especially demanding in relation to

clinical applications. Legal requirements that apply to patient-related health care data and the broader regulatory environment in which health care is provided are substantial, complex, and often difficult to navigate, both in the United States and Europe. Accordingly, ensuring that blockchain-enabled *clinical* applications demonstrably comply with applicable laws will be very challenging [9,92]. That said, blockchain advocates place considerable optimism in *smart contracts*—computer programs that automate the verification, execution, and enforcement of certain terms and conditions of an arrangement, built on top of a blockchain system that enables a distributed ledger to function as a distributed computer [8,93-95]. Accordingly, blockchain systems in combination with smart contracts could operate as a form of *RegTech* through which compliance with legal and other regulatory requirements can be *hardwired* to execute automatically. However, whether these technologies will live up to this potential, given the messiness and unpredictability of real-world contexts, remains unknown [96,97]. Overcoming each of these barriers to organizational commitment presents formidable challenges and taken together, generate major obstacles that lie in the path of successful blockchain innovation and diffusion in health care [98].

Interoperability and Standardization

Another significant obstacle that continues to frustrate health care data-sharing efforts is the lack of interoperability within and between health IT systems [99]. This is equally true for blockchain-based data management systems. HIMSS explains that, for the purposes of a health data ecosystem, interoperability is "the ability of health information systems to work together within and across organizational boundaries in order to advance the effective delivery of healthcare for individuals and communities." HIMSS [100] notes that:

> *In the healthcare space, interoperability is important across systems and organizations for information to flow seamlessly between actors like patients, doctors, hospitals, payers, etc. However, it has been one of the hardest challenges healthcare has faced as vendors, providers, policies, payers and patients have, at times, set roadblocks up and created misaligned incentives to achieve true exchange of data between disparate systems. Realizing the benefits of blockchain technology depends on the associated network of healthcare organizations being able to share data and collaborate via the blockchain. These net-new requirements increase the scope of interoperability challenges that pre-dated blockchain. When systems are not interoperable, the cost of verification and networking of the information is very high. The same is true for the cost of settlement and reconciliation of the transactions of information.*

To maximize the capacity for data sharing, *full* interoperability between blockchain systems is needed; however, achieving this remains a very long way off and arguably will forever remain an impossible dream [101]. Interoperability in relation to health care IT can be understood at 3 levels. First, foundational interoperability allows data exchange from one IT system to be received by another without needing the receiving IT system to have the capacity to interpret that data. Second, structural interoperability allows the movement of health care data from one IT system to another such that the clinical or operational purpose and meaning of the data are preserved and unaltered, which is enabled by defining the structure or format of data exchange. Third, semantic interoperability refers to the ability of 2 or more IT systems or elements to exchange information and use the information that has been exchanged [102], requiring both the structuring of the data exchange and the codification of the data (including vocabulary) so that the receiving IT systems can interpret the data. Critical to the achievement of interoperability is the shared use of common standards (that are either mandated or very widely accepted and adopted) at the relevant level. To this end, important and ongoing efforts are being made to establish international standards for health care data and IT architecture to enable all 3 levels of interoperability for health care data sharing, including the work of standard-setting organizations such as Integrating Healthcare Enterprise International [103] and Health Level Seven International [104,105].

The need for interoperability may help explain why the blockchain use cases with the greatest prospects of success in the near term are largely concerned with sharing *back office* functions and associated records, such as credentialing and provider identity verification, and claims management and billing rather than the sharing of patient data for clinical purposes [106,107]. Although these applications entail interorganizational record sharing, which may well be in different formats and structures, the substantive content of the records themselves is likely to have a high degree of *semantic* interoperability that tends to avoid subjective evaluations and is not readily prone to misinterpretation or misunderstanding. For example, the ProCredEx consortia enable the pooling of records concerning whether a particular individual possesses particular professional qualifications; these records can be readily shared between participants without any serious likelihood that the meaning of those records will be misinterpreted by other members [57].

In other words, blockchain-based record sharing among multiple participants is more likely to succeed if the underlying data are highly stable and readily verifiable as evidencing the *truth* of the underlying phenomena that the data purports to represent (eg, such as whether a clinician has obtained a university degree from an approved medical school), and the consequences of an error or data inaccuracy are not safety-critical. However, experts worry that, as multiple blockchains for health care networks emerge supporting a wide variety of applications, they might not be interoperable *with each other*. This could seriously undermine the value and benefits of blockchain-enabled solutions [108,109]. For example, an organization participating in a pharmaceutical supply-chain blockchain may find that it is not interoperable with a clinical trial blockchain network that it also wishes to join, significantly reducing the potential value it might otherwise derive from using blockchain. These anticipated challenges reflect the observations by Greenhalgh et al [98] that a common but significant health care infrastructure challenge in many countries lies in enabling state-of-the-art

individual technologies to interface with (but have often been designed with little awareness of) a legacy infrastructure and restrictive regulatory standards, all in the context of a complex, fast-changing, unpredictable, and underfunded service environment.

Internal Governance

A feature of the growing number of health care consortia seeking to use blockchain technologies to selectively share data and collaborate at the B2B level is that member organizations are often conventionally regarded as competitors [110]. If these novel sociotechnical collaborations are to succeed, a clear set of agreed norms and arrangements to govern the terms of their collaboration will be essential, without which stable cooperation across the network is unlikely to be viable in practice. In open, permissionless blockchains, including Bitcoin and Ethereum, disagreements over proposed changes to the technical architecture have led to highly publicized disagreements, reflecting the different political viewpoints and motivations of participants, which often lead to the fragmentation of the blockchain network in the form of a *fork* in the ledger [111-114]. Although the overwhelming majority of health care distributed ledger technology applications use hybrid enterprise or permissioned systems, they face internal governance challenges that may be no less fraught than those that have arisen in open, permissionless blockchains [115]. However, establishing a durable internal governance framework that commands widespread acceptance by members and participants across the network (including potential new members and participants) will be especially challenging, which is discussed more fully in the following sections in relation to the challenges of *coopetition* [116].

Data Security and Integrity

Although the most promising engagements with blockchain in health care contexts have hitherto largely avoided the clinic, many industry experts believe that as the technology matures, clinical applications entailing collaboration over patient data are likely to emerge [117]. A further 3 challenges are already being encountered within the health care sector in seeking to engage with blockchain technologies, but which are likely to be particularly acute if blockchain applications are applied in clinical contexts. Chief among them are the challenges associated with data security and integrity. For any blockchain-based system, the desired level of data security must be identified, established, and maintained across the system's architecture and operation. For patient data, questions about data security are primarily informed by their size and high sensitivity. Accordingly, it is widely accepted that patient data are best stored *off chain*, for example, in a relational database with the shared ledger merely storing metadata together with pointers to where the actual patient data resides and hash codes to verify the integrity of the off-chain data [82,118]. By incorporating technological mechanisms for identity management and access control into blockchain systems, the on-chain record can also record when, whether, and by whom the relevant linked data are accessed. This approach is reflected in what HIMSS refers to as the principle of *minimal sufficiency* to on-chain data, which it advocates as best practice, stating the following [92]:

> *Blockchain technology was not designed to be a storage mechanism for data and should not be leveraged as such, for security, privacy, compliance and performance reasons. Information added to the chain may be transparent to permitted network participants and difficult to remove without affecting the entire chain. Therefore, we strongly recommend that regulators and policymakers promote that organizations leveraging blockchain-enabled solutions employ a "minimal but sufficient" strategy for the data that should be included on-chain. This strategy should be guided by the use case's data needs, and implementers should keep in mind not only the privacy and security risks but also the performance of blockchain transactions when deciding the amount of data and/or personally identifiable information (PII) included on the chain. Whenever possible, privacy-enhancing technologies should be used to secure private data on the blockchain.*

However, by storing patient data offline, the blockchain ledger cannot ensure data security. Although there are various measures, including encryption, that can enhance off-chain data security, these are separate and distinct from blockchain's technological protections for securing on-chain data. Given that one of blockchain's most significant *promises* lies in the iron-clad security it purportedly brings to on-chain data, the inability to extend that protection to the data stored off chain seriously limits the technology's capacity to deliver on this promise. In addition, a frequently overlooked and neglected issue in the blockchain literature is the problem of *data leakage* or *escape*. Although blockchain offers the technological capacity for fine-grained, auditable data sharing through technological access controls that protect the privacy and confidentiality of stored records, blockchains do not address the possibility that, once data are revealed, those with access will generally be able to copy and extract the data and store it perpetually. As Finck [93] warns, the purposes of blockchain projects can be completely undermined when data escape is possible, particularly in circumstances where data are sold once, and the buyer can then resell or manipulate the data set at will.

Truth and Immutability

A critical feature emphasized by blockchain advocates is the technology's capacity to create a single, immutable, shared authoritative record, reflected in descriptions of blockchain as a *truth machine*. However, the validity of this claim assumes that the data recorded on the blockchain has *integrity*—meaning that they are accurate, up to date, and comprehensive. Although often discussed in terms of providing assurance that the data have not been tampered with or subject to unauthorized alteration [119,120], data integrity also requires that the data faithfully and accurately represent the underlying real-world phenomenon they purport to represent. For example, if a health care record is labeled as representing a particular item, such as an X-ray taken of patient Y on Z date at location P and stored

at an off-chain location Q, then it must represent precisely that. This also requires (among other things) that blockchain consistently keeps transaction information associated with the correct person. Although considerable work is being devoted toward identifying universal identity management solutions for blockchain-based systems, this remains an unresolved challenge. Questions concerning how blockchain can ensure that the records appended to the blockchain ledger are accurately tethered to and reflect the underlying reality that they purport to be associated with remain surprisingly overlooked (or perhaps conveniently ignored) in the existing blockchain literature. In health care, these challenges are particularly acute if data are to have clinical relevance. However, they sit uncomfortably with developers' and entrepreneurs' claims that blockchain will enable the creation of accurate and comprehensive patient records that draw data directly from a patient's wearable device, which is gathered via sensor technologies, without mechanisms that guarantee the accuracy, veracity, and reliability of data thereby collected [21]. Although sensor technologies can help establish direct connections between the digital and physical world in a secure manner, sensors can be easily tricked (even if tamper-proof hardware is used); thus, conventional social trust is needed to trust the veracity of the underlying data. This calls into question the value that blockchain purports to offer in the first place. In the context of supply chain management, Wüst and Gervais [119] refer to the following:

> *The inherent problem of the interface between the digital and the physical world. A human, or some machine under the control of a single writer, typically is required to register that a good has arrived in a warehouse and, if for example, its quality is appropriate. If there is no trust in the operation of these employees, then the whole supply chain is technically compromised as any data can be supplied by a malicious writer. If, on the other hand, all writers are trusted, a blockchain is not needed as a regular database with shared write access can be used instead.*

Although the accuracy of health care administrative data may not always be safety-critical, there is little value in a blockchain that maintains a tamper-proof record of inaccurate or poor-quality data. Accordingly, descriptions of blockchain as a *truth machine* appear somewhat overstated [31]. Data integrity is a necessary prerequisite for establishing trustworthiness and providing a reliable basis for decision-making. Houlder [121] explains the following:

> *Data security is often equated with protecting data confidentiality: but it is data integrity that must be protected – and this requires ensuring that the data is accurate, up-to-date and complete. Blockchain only protects what arrives at blockchain. From an immutability and transparency standpoint, if that data is compromised or of poor quality before it reaches the blockchain you end up with garbage in, garbage out, where you're protecting garbage on the blockchain. Unless the data has the integrity and quality needed, blockchain will not realise its potential.*

The yet-to-be-resolved challenge of assuring data integrity may help explain why the handful of patient-facing blockchain applications offered through the NHS (in partnership with private technology providers) have largely confined blockchain functionality to storing records of patient consent and providing computational assurance concerning when a patient has accessed (and securely copied) his or her NHS records.

In addition, the application of blockchain to health care will entail human interaction and engagement. Even if users are provided with high-quality training, help, and guidance, mistakes will be inevitable. However, identifying and correcting errors generates new challenges, particularly given that blockchain ledgers cannot be retroactively altered. Although blockchain's tamper-proof character prevents certain kinds of mistakes and problems that arise in relation to conventional centralized databases (such as inadvertently overwriting data stored on the database), they will inevitably introduce new ones. This is particularly so given the complex and dynamic nature of health care settings and the multiple intersecting and sometimes conflicting interests, rights and obligations, expectations, and anxieties they typically implicate. In other words, the vagaries of human behavior and decision-making in real-world health care contexts could prevent blockchain's expected benefits from being fully realized.

Quality and Safety

Data integrity is but one element of the larger, multidimensional challenge of ensuring that blockchain systems can be implemented in health care without compromising quality and safety, particularly when they interface with and operate within clinical contexts. Although the safety implications arising from the digital transformation of health care have been well-documented [122], patient safety has received relatively little attention to date, perhaps because of the relative scarcity of clinical blockchain use cases being trialed and implemented. Patient safety issues include the need to ensure that blockchain systems that affect clinical workflows and environments are subject to rigorous testing, validation, and independent evaluation before their introduction into the clinic. However, in the US health care context, no systematic attention appears to have been given to the technical robustness, safety, or ethical dimensions of blockchain implementation. Rather, health care blockchain projects appear to move from proof of concept through to implementation without necessarily being subjected to formal testing and validation to provide assurance of the system's robustness, resilience, or clinical safety. Accordingly, health care blockchain expert Heather Flannery argues that these projects should be understood as experiments to learn and discover the kinds of conditions that produce the desired results; although these conditions are partially technical, in her view, most are clinical, social, economic, legal, ethical, and governance-related. Hence, Flannery argues that these projects should be undertaken as studies subject to proper research protocols as the only ethically viable way to move past the point of experimenting with the technology [123,124]. In contrast, regulatory oversight of health IT systems in the United Kingdom is more developed. Thus, the Health and Social Care Act 2012 requires the development and implementation of clinical risk management processes to ensure patient safety with respect to

the manufacture of new health IT systems, or the modification or decommissioning of an existing system, and their deployment and use [125,126].

Normative Conflict Requiring Satisfactory Resolution

Although the above 6 challenges might be overcome in time, several further challenges rooted in inherent tensions between desirable qualities or functional requirements that arise when configuring blockchain applications may prove more intractable. Although technical blockchain experts have identified various *design trade-offs* when building blockchain systems for specific purposes [82], reflecting the *micropolitical nature of software choices* [98], the following 3 are particularly apposite to health care settings:

Performance and Scalability Versus Security

Ideally, an IT system would provide high levels of security and resilience for the system and the data it stores and generates while offering high-speed performance at scale. However, achieving these objectives in blockchain systems is technically impossible (on currently available technology) because of inherent trade-offs in functionality. There is a loose *trilemma* in the design and configuration of blockchain systems such that they can have at most two of the following three properties: (1) decentralization, (2) scalability, and (3) security. The more computationally demanding and time intensive the consensus protocols for validating transactions [127] before being appended to the ledger and the greater the degree of decentralization across the network, the greater the ledger's security and resilience against attack but the slower the performance in terms of speed and throughput [128]. In addition, as blockchains ledgers are distributed, they entail high storage costs, precluding large data from being effectively stored on the blockchain. As already noted, although blockchain can be used for access control (and auditing), large data must be stored *off chain* [129]. This generates the need to assure off-chain data security; simply using a blockchain to manage access does not thereby offer the security of the stored data, pointing to an inherent tension between the desired values of scalability versus security in their design and operation [130].

Transparency Versus Privacy

The transparency of all transactions appended to the ledger is a critical feature of open, permissionless blockchains, contributing to the ledger's trustworthiness. However, this level of transparency is fundamentally at odds with the private and confidential nature of patient data and other health-related personal information. Accordingly, there is an inherent tension between the need to respect privacy and confidentiality and the design and operation of open blockchain networks that are transparent to the world at large [93]. Under the European Union's General Data Protection Regulation, data collectors and processors must not collect or process personal data without a lawful basis, which includes, but is not limited to, consent by the data subject. In common law systems, additional legal obligations attach to information acquired in circumstances of confidence by clinicians and other care providers, including information communicated by patients to doctors during clinical encounters. At the same time, health care organizations typically seek to keep details of their business records confidential. As already indicated, the *minimum sufficiency* principle of information collection and storage is considered best practice in health care IT design and operation so that health care blockchain systems have hitherto largely taken the form of *permissioned* systems, thereby limiting *read* access to the ledger to those authorized to do so, with health care data being stored off chain and only the hashed metadata stored on chain.

Computational Trust Versus Social Trust

Although the hyperbole associated with blockchain technologies has begun to subside, early industry activity and engagement with blockchain appears largely to have been *technology lead*, arguably motivated by the desire to engage with the latest cutting-edge technology [131] without either a deep understanding of the technology or a desire to meet a specific need that blockchain might usefully address. As the technology has begun to mature, so too have academic investigations, with more recent academic critiques arguing that conventional database structures and systems will often provide cheaper, faster, more sustainable, and scalable approaches to specific real-world data access, management, and storage needs [119,132]. Several scholars have sought to identify the circumstances in which blockchain technologies have a genuine prospect of addressing real-world problems for which conventional centralized, networked databases are inadequate. For example, Wüst and Gervais [119] argue that if either a trusted third party who can verify transactions is always available on the web or if all writers mutually trust others, then a conventional database with shared write access is likely to be preferable. However, if the writers of the ledger do not trust each other, a permissioned blockchain may provide a viable solution. Their insight resonates with, and may help to explain, the recent emergence of consortia of health care organizations who compete with each other in many respects but recognize that they share a limited common interest in relation to specific kinds of data and forms of data sharing. In these circumstances, member organizations may not fully trust each other, and a permissioned blockchain may provide a technological solution to address this trust deficit.

Permissioned systems (including so-called *enterprise* systems) have hitherto been favored in health care blockchain applications. They appear to offer the greatest opportunities for generating value from network effects arising from the creation of a shared, trustworthy transactions ledger between a limited number of authorized participants, enabling members to benefit from effectively sharing elements of a *back office* administrative system (ie, shared records management in a single shared ledger). However, the security and integrity of the shared record will only be as strong as its weakest link, giving rise to what may be called the *computational trust* paradox. On the one hand, the use of computational consensus mechanisms to verify transactions and maintain a single shared ledger across a distributed network of computers obviates the need for each participant to maintain their own ledger and reconcile their ledger with that of other participants, generating significant administrative efficiencies. On the other hand, this means that each participant must trust the integrity, quality, and security of the participant's records, including the accuracy of the

information contained in those records. This relies, in turn, upon the integrity of the practices and processes underpinning the events and transactions that each record purports to represent (subject to those records being validated by the network and appended to the shared ledger). By pooling their records, each participant is thereby exposed to the vulnerabilities of the record keeping and information security practices of their fellow participants. A paradox arises because, although blockchain systems offer the possibility of coordination with others *without* the need for social trust in those others by relying instead on *computational trust*, when these systems take the form of permissioned ledgers, this is likely to *enhance* the need for conventional (social) trust in the practices and systems of other network members. Hence, reliance on computational trust serves, ultimately, to heighten the importance and need for social trust to reap the gains from cooperation [133]. A leading industry expert [121] commented as follows:

> *If you have a consortium of 10 organisations and one of them is a kind of weak link...in terms of they're connecting to the blockchain and pulling information off the blockchain, or in interactions that are enabled by blockchain, and putting information in their own enterprise systems which are insecure and a breach occurs, it impacts everyone...you're not just worried about securing your own organisation, you've got to make sure that everyone connecting to that blockchain is adequately secure.*

Discussion

Overview

The above findings draw attention to the number and variety of formidable, multidimensional challenges that must be satisfactorily addressed if blockchain technologies are to traverse the valley of death. Nevertheless, this study also shows how early engagements with blockchain in health care have begun to demonstrate that these technologies *can* enable and facilitate novel, valuable forms of health care data sharing and cooperation in real-world health care settings under specific circumstances. The following discussion critically reflects upon these key findings by reference to 3 quite different strands of academic literature: first, it examines the character and tensions inherent in these incipient forms of blockchain-enabled cooperation; second, it draws on insights from innovation studies to consider the character of blockchain as *radical* or *incremental* health care technologies; and third, it evaluates the prospects of blockchain for health care's *favorite use case*, that is, to enable *patient sovereignty* over EHRs while overcoming the currently siloed approach to patient data.

The Blockchain in Health Care Promise: Rhetoric, Reality, and Blockchain-Enabled Coopetition

Recent health care blockchain initiatives have successfully demonstrated how technology can facilitate limited, purpose-specific forms of cooperation through the shared pooling of data to solve common problems. These incipient forms of cooperation involving data sharing, which consortia of health care organizations are actively seeking to develop, rely upon interactions that organizational studies scholars refer to as *coopetition*. Although the interaction between firms conventionally focuses on either of 2 contrasting and antagonistic logics—competition (traditionally defined as the conflicting and rivalrous relationship between firms) on the one hand and cooperation on the other [134]—the actual interaction between firms can involve cooperation in some activities while competing in others [135]. As these coopetitive relationships entail 2 diametrically opposing logics of interaction, they are inherently complex, generating tensions between and within organizations [136,137]. In particular, each organization faces conflicting incentives: a desire to cooperate through the pooling of resources to generate shared benefits on the one hand and the incentive to capture private benefits for themselves on the other [137,138]. Several theories have been offered, claiming to predict the effect of coopetition on firm performance [137,139], including transaction cost theories. Transaction cost theories highlight the difficulties of fostering trust between parties in a coopetitive relationship, as each participant has incentives to opportunistically behave while facing high levels of uncertainty [140]. This creates barriers to open collaboration, making it difficult to develop the necessary level of trust needed for common projects to succeed. On the other hand, several scholars have noted that the inherent tension between cooperation and competition arising within coopetitive relationships can, if managed efficiently, create shared value and generate beneficial outcomes for each party [137]. On this view, the key to successful coopetition is effective management of the coopetitive tensions within the relationship that nurtures and maintains an appropriate balance between competition and collaboration. Various insights from this literature help illuminate the challenges likely to arise within health care consortia comprising organizations that would otherwise regard themselves as competitors seeking to share health care data via a permissioned blockchain network. The configuration of access rules within a permissioned blockchain system offers a potentially novel mechanism through which *social distrust* might be replaced by *computational trust*, thereby enabling new forms of cooperation at the B2B level that might not otherwise occur, as the trust gap is too difficult to overcome.

That said, the need for strategies to successfully manage the tension between organizations within the network highlights the importance of a blockchain network's internal governance arrangements, particularly in establishing and maintaining rules concerning changes to the network's architecture, protocols, rules of admission, and incentive mechanisms. Details about the internal governance arrangements that underpin the various blockchain consortia that have emerged in recent years are not publicly available because of the confidential nature of the underlying partnership agreements. Nevertheless, one would expect that formulating and maintaining a set of governance arrangements that will prove durable and resilient over time is likely to be extremely difficult, as the interests of participants are not *wholly* aligned and hence, difficult to settle in advance. These internal governance arrangements must identify how critical decisions about, for example, the structure of economic incentives or rewards for participants to engage in the consensus protocol, how the network will be maintained, and how other changes to the network's structure and operation will be made and by whom. Potentially conflicting interests may not be

readily apparent at the time of the network's inception but may subsequently arise as organizational needs, the larger industry context, and other external factors inevitably change over time. However, perhaps it is only when the *rubber hits the road* in the context of specific disagreement between network participants that the critical need for robust governance structures that command the respect and allegiance of the network's members becomes apparent.

Finding an appropriate but flexible equilibrium between the need for social trust between network participants on the one hand and *hardwiring* computational trust into the technical architecture and operation of the blockchain system will be a very tough nut to crack. It remains an ongoing and continuing challenge for the foreseeable future, at least until blockchain in health care achieves a level of maturity in which a widely shared set of agreed norms concerning the core requirements of good blockchain governance can emerge. Furthermore, these networks are likely to encounter what has been referred to above as the *computational trust* paradox; by seeking to rely on computational trust to provide a basis for enabling cooperation between firms, this may reinforce and accentuate the need to nurture and maintain social trust between the parties in domains in which computational trust cannot be secured.

Permissioned Blockchains in Health Care: Incremental or Radical Innovation?

Early engagement with blockchain technologies in health care contexts has largely used *permissioned* blockchain systems (including so-called *enterprise* blockchains) rather than open, permissionless systems that many blockchain advocates believe will revolutionize social cooperation between strangers without the need for conventional trusted third-party intermediaries, such as banks and governments. Although the blockchain for health care promise has not typically emphasized the role of permissionless blockchains, a significant *revolutionary* narrative is evident in claims that blockchain technology will *disrupt* health care [17]. Within this narrative, blockchain is portrayed as enabling radical patient empowerment achieved via hard-coded access controls, securely facilitating the seamless exchange of patient records currently stored in disparate data silos while enabling patients to exert fine-grained control over who is granted access to those records, wherever located, and on what terms [141,142]. It is blockchain's potential to overcome the many existing technical, legal, and bureaucratic obstacles that obstruct the free flow of health data while preventing patients from exerting control over access to their own health care records that is typically highlighted rather than emphasizing the role of permissionless blockchains in achieving this vision.

The health care sector's focus on permissioned (rather than permissionless) blockchain systems to date will disappoint radical blockchain enthusiasts such as Bruce Schneier, a well-known computer security expert. Schneier claims that permissioned blockchains are "completely uninteresting" as they are no different from centralized databases in relying on conventional forms of trust to facilitate social cooperation [143]. However, his perspective fails to recognize that even open, permissionless blockchains ultimately and invariably rely on social trust to support their internal governance arrangements [112,113]. By enabling participants to rely on computational trust mechanisms to enforce the explicitly agreed terms of their cooperation, permissioned blockchains could facilitate novel forms of cooperation, leading to valuable and potentially transformational change.

The transformational potential of permissioned blockchain systems can be illuminated by reference to a distinction drawn by technology management researchers between *inventions*, which refers to an "idea, sketch, or model for a new or improved device, product, process or system" and *innovations*, which, in the economic sense, arise "only with the first commercial transaction involving the new product, system or device" [144]. Within this literature, a further distinction is commonly made between *incremental* and *radical* innovation. Incremental innovations introduce relatively minor changes to an existing product, exploiting the potential of the established design (although they may entail considerable skill and ingenuity) [145-148]. In contrast, radical innovations are based on different engineering and scientific principles, potentially opening up whole new markets and applications [146,147,149]. This distinction is considered to have important competitive consequences: incremental innovation tends to *reinforce* the capabilities of established organizations, whereas radical innovation forces them to *ask a new set of questions*, draw on new technical and commercial skills, and use new problem-solving approaches [36,146,148,150]. Hence, radical innovation often creates great difficulties for established firms [148,151-153]. As the late Clayton Christensen famously argued, radical innovation can *disrupt* the entire industry as incumbent firms lose out to smaller firms with fewer resources who use these radical innovations to deliver superior functional performance, eventually leading to their adoption by the incumbent's established customers [154].

It is too early to assess whether permissionless blockchains will ultimately form the basic technological architecture underpinning a dominant design for particular kinds of health care applications. It is not possible to reliably predict whether they will come to be regarded as incremental or radical innovations or, indeed, whether they come to be understood as innovations at all or merely just an interesting invention with little real-world utility. However, the preceding insights suggest that we cannot assume that the use of permissioned rather than permissionless blockchains in health care necessarily implies that they cannot subsequently establish themselves as *radical innovations*. Bruce Schneier's dismissal of permissioned blockchains overlooks the possibility that they may offer particular functional properties capable of meeting a very specific sectoral need. In functional terms, permissioned blockchains are something of a hybrid—analogous to centralized databases insofar as participation and access require authorization from a network controller but can mathematically provide the verified security and automated audit functionality of permissionless blockchains without the heavy computational expense. Accordingly, they might prove *radical* in their effects [155], creating shared value by facilitating new forms of coopetitive collaboration between health care organizations while generating significant efficiencies.

Blockchain-Enabled EHRs and the Promise of Patient Data Sovereignty

Mention has already been made of frequent references in both academic and industry literature to the potential of blockchain-enabled EHRs to *revolutionize* health care. According to this vision, blockchain operates as a form of *middleware*, facilitating the flow of data between independent IT systems, thereby integrating disparately located silos of patient health care records while enabling patients to exercise fine-grained but automated control over their EHRs via *smart contracts* that automate the execution of patient instructions or consents [14,109]. In so doing, blockchain is widely portrayed as *solving* the tension between the need to respect the privacy and confidentiality of individual health records and recognizing the tremendous knowledge potential unleashed by pooling and aggregating patient data. Blockchain for health care scholarship has largely taken these claims at face value. Although various blockchain-enabled EHR projects are currently being experimented with, the preponderance of activity appears to reflect the views of industry experts, who advise that blockchain applications focusing on patient care records are best avoided in the early stages of the sector's engagement with the technology because of the demanding and complex legal and regulatory requirements that apply to personal health information. Nevertheless, many believe that these demands can eventually be overcome, so that patient-controlled, blockchain-enabled EHRs will eventually emerge [91,117]. Even if the difficult and complex challenges and tensions identified in the preceding section can be satisfactorily overcome, whether these systems will live up to their promise depends upon the validity of underlying assumptions about their benefits, particularly in clinical contexts.

Academic literature concerned with health IT implementation has hitherto been dominated by the *technology acceptance model* (TAM), which originates from the perspective of health information systems research that studies the systems that support health and medical work [156]. This approach, developed jointly by doctors with an interest in computers and computer scientists with an interest in medicine, adopts a positivist outlook that regards the relationship between technology and their expected benefit as linear and causal [81]. Similar assumptions appear to underpin 2 distinct, albeit related, elements of the vision of blockchain-enabled EHRs offered by their proponents. The first assumption is that blockchain will enable the seamless integration of patient records currently stored in distinct *data silos*, thereby effectively providing patients and clinicians with a single, up-to-date, and comprehensive repository of each patient's medical records over his or her lifetime. The second assumption is that configuring blockchain-enabled EHR systems to enable patients to control access to their records will ensure respect for privacy without impeding the transfer and sharing of patient data for clinical and other legitimate purposes.

In relation to the first assumption, Berg et al [157] highlighted studies of EHR implementation that emphasize the personalized nature of health care work, thus setting "natural limits to the possibilities of IT to revolutionise this work." These studies suggest that the considerable energy and resources thus far devoted to various components of the EHR failed to substantially improve the quality or efficiency of frontline clinical work. Moreover, they argue that EHRs are *unlikely ever* to produce dramatic gains in the quality of care because of the importance of the creative human work needed to bridge the gap between clinical design and technical reality [50]. In contrast, biomedical literature often rests on a TAM model, evoking a belief that IT systems will make clinical information instantly available, implicitly assuming that *meaning* can be transmitted unproblematically along with the data contained in the EHR. This assumption also appears to underpin the *beautiful dream* of blockchain-enabled EHR systems, suggesting that, if properly designed, blockchain will enable more accurate, secure, and timely data access that will drive quality improvements across health care, including clinical care. However, studies from more sociologically informed traditions roundly reject this assumption, providing evidence that clinical data must be interpreted in context and *framed* before they become meaningful, and this is not achieved simply by placing information on an electronic platform that is accessible by multiple users [50]. These studies rest on assumptions that differ from those underpinning the TAM model (including understandings of what counts as *success* for the purposes of evaluating health IT systems), drawing on a wider range of methodological approaches [158]. They highlight the critical role of context, culture, and the values brought to bear by medical professionals in clinical environments, identifying the clinical consultation as a complex social encounter that occurs within a heavily institutionalized environment [50]. In particular, scholars from these traditions emphasize the role of values in clinicians' understanding of what constitutes and how they seek to practice *excellent care* and the nature of clinical knowledge as *tacit, context-bound, and ephemeral rather than codifiable, transferable and enduring* [50]. These studies find that in many *failed* EHR projects, technical designers typically missed these subtleties and produced artifacts that fitted poorly with the situated nature of knowledge and the microdetail of clinical work practices [81].

Conversely, the TAM model assumes that the failure of EHRs to generate their expected benefits can be largely explained by shortcomings in EHR system design and implementation, resulting in user *resistance* to the technology [50], which can be overcome by system design improvements and more sensitive implementation. One of the most significant lessons from studies of EHR adoption is that considerable clinician resistance to these systems has often resulted from the unintended but radical alteration and disruption of clinical workflows and patient interactions accompanying the introduction of EHRs, adversely affecting physician workload [81,122,159]. In addition, the data captured in these records have been described by a legal expert as often error prone, incomplete, unprotected, and dispersed across numerous organizations unknown to the patient [130]. However, if used as *middleware* to support the integration of EHRs across disparate data silos, blockchain technologies should not, in theory, affect clinical workflows or practices, as they operate as a *backend* technology. In other words, blockchain should not exacerbate workload burdens or introduce additional integration work of the kind that has typically accompanied EHR implementation. However, HIMSS [160] warns that:

> *Blockchain technology is meant to operate on the back end and should avoid, where possible, adding additional steps for the end user. If this technology is poorly implemented without a strong federated approach, workflow may be negatively impacted, potentially affecting the quality of care.*
>
> *Adding a blockchain-enabled solution without considering its applicability to a use case, potential value, or relationship with a partnering stakeholder, could add a layer of complexity to an already convoluted health IT ecosystem. Furthermore, if this complexity detracts from the end goal of sound care, ethical and other questions may arise from the implementation of this kind of solution that is not well suited for the coordination and delivery of care.*

This cautious view, offered by one of the health care IT sectors' leading industry bodies, suggests that one cannot assume that implementation of blockchain as a *backend* technology will not add further system complexity simply by avoiding any change to the *front end* experience of users [98].

The second assumption underpinning the promise of blockchain-enabled EHR systems—that they will enable patient data sovereignty and empowerment by providing patients with fine-grained control over who may securely access their health care records and on what terms—resonates with broader societal shifts in favor of *self-service economy* models across many sectors [161]. Blockchain for health care advocates assumes that patient-controlled health care records will resolve dilemmas arising from the tension between ensuring compliance with privacy and confidentiality norms and duties and fostering the aggregation of patient care records for medical research and other legitimate purposes [162]. This assumption is likely to be both mistaken and naive for several reasons. First, although it is essential that patients are enabled to exercise their rights of informational self-determination in relation to health and care information that directly pertains to them [163], some patients will enthusiastically embrace the capacity to control access to their medical records [164], whereas others will not [165]. A significant proportion of patients (including children and those with mental illness or disability) cannot make informed decisions about access to and sharing of their clinical data; therefore, alternative consent models will be needed. Second, there will be circumstances in which patient consent cannot be secured; however, the ability to access that patient's health records will be necessary, particularly in emergency situations. Third, in other circumstances, patient rights to data privacy and confidentiality may be legitimately overridden by compelling public interests, particularly when appropriate technical and organizational measures are taken to safeguard those interests so that patient consent is neither legally nor ethically needed to justify access to their records. For example, the public interest in data sharing and access can, particularly in times of public health care emergency, justifiably override individual rights to privacy and the need for patient consent to data sharing.

Taken together, the vision of *patient data sovereignty* underpinning blockchain-enabled health records exhibits a *techno-solutionist* mindset, reflecting a belief that technological fixes can be used to *solve* complex social problems [166]. Critiques of this kind draw attention to the complex and messy set of interacting norms, practices, and dynamics that arise in the real-world contexts for which simple technological *solutions* are unlikely to be found, which the disappointing experience of EHR implementation has vividly demonstrated. Despite this, the underlying *information as property* paradigm that informs these beliefs appears to hold considerable sway in contemporary debates about data governance, both within health care and beyond [167]. However, as the German Data Ethics Commission points out, supporting and promoting the practical capacity of patients to access their health care records and take an active role in decisions about access to those records does not necessitate a *propertisation* model of data governance, which underpins the views of those who favor the making of micropayments to data subjects in return for access to their personal data [163,168,169].

Apart from unfounded assumptions that technological access controls built into health IT systems can be expected to successfully mediate and resolve the complex interests in patient data, any such shift in favor of patient sovereignty over health care records may reflect an underlying reconfiguration of the ethical basis upon which patients receive care. Rather than characterizing patients as vulnerable individuals and beneficiaries of care provided by an expert clinician who is professionally obliged to act in their best interests, patients may increasingly be expected to *self-manage* their own medical records based on largely unexamined claims that this will *empower* them to exert greater control over their own care [85]. However, if these systems are configured in ways that force adherence to norms that are at odds with the basic cultural and moral norms of contemporary health care settings, this might provoke backlash. This might occur, for example, if systems are designed to shift responsibility for care to patients and away from clinicians and care providers in ways considered contrary to the professional duty and commitment of clinicians to act in the best interests of their patients [170]. Alternatively, if patients are unable to make meaningful decisions about the sharing of their data, they are likely to stick with defaults; however, these might fail to reflect legitimate patient expectations and existing informational privacy and confidentiality norms. These and other similar risks underline the importance of meaningful stakeholder engagement in the design and implementation of blockchain systems, as well as the importance of more critical engagement with the logic of *empowerment* that frequently accompanies their promised benefits [50].

Accordingly, the vision underpinning blockchain for health care's *favorite use case*, which is capable of seamlessly integrating patient-controlled EHRs accessible by clinicians in real time wherever located, is based on problematic and unrealistic assumptions. Their promised benefits are therefore unlikely to be realized in practice. Even if blockchain technologies can operate as *middleware*, enabling organizations to access and retrieve patient health care records stored at different sites in real time, blockchain cannot ensure that the information stored in those records is an accurate representation of the underlying real-world events they purport to reflect. Conferring responsibility for making access decisions on patients themselves will not avoid conflicting and sometimes difficult

trade-offs between respect for patient privacy and other rights and legitimate interests implicated in health care. In other words, the health care *data silo* problem is only one aspect of a much larger, more complex set of needs, rights, and interests associated with medical and health care information, which are likely to continue to defy hard-coded solutions.

However, one of the most valuable insights arising from *technology-in-practice* studies of EHRs for understanding the real value that blockchain-enabled EHRs might create arises from the recognition that there are two conflicting work processes in play with EHRs: immediate clinical care (primary use) on the one hand and tasks such as audit and research, which are one step removed from the clinical encounter (secondary use) on the other. These studies show that when EHRs are used as a formal tool (eg, with structured templates and a requirement for data to be coded), they often slow down and frustrate the clinical encounter but can greatly accelerate secondary uses of clinical data. Accordingly, Greenhalgh et al [50] argue that, rather than promising that the EHR will *save time* or *make clinical care more efficient*, a more honest message would be that creating accurate and complete clinical records requires the sacrifice of time and effort by frontline clinical and admin staff, which is (sometimes) justified by more benefits for efficient business processes (eg, billing) governance and research. In other words, the real value of blockchain-enabled health records might rest in their secondary use, enhancing health care administration and medical research rather than in substantially improving clinical care.

Study Strengths and Limitations

This study has both empirical and theoretical strengths. Its strengths are 3-fold. First, by focusing attention on the health care sector's experience in attempting to use blockchain technologies for specific purposes, it fills an important gap in the existing literature by offering a more grounded, evidence-based but theoretically informed appraisal of blockchain's prospects for transforming health care, unlike the burgeoning academic literature that has focused overwhelmingly on the technology's *potential* to solve health care problems. Second, it identifies and critically examines a series of complex, multidimensional implementation challenges that must be satisfactorily confronted if blockchain technologies are to be widely taken up within health care contexts, illuminating the nature and dimension of these difficulties by referring to a range of theoretical perspectives. Third, the critique offered here contrasts sharply with the considerable hype that accompanied the emergence of blockchain technologies. It provides a more theoretically and empirically grounded appraisal of the oft-proclaimed radically transformative potential of blockchain technologies in a specific domain, notably within complex health care environments that are typically saturated with technologies, many of which are highly sophisticated and often safety-critical.

The limitations of the study rest in the lack of publicly available, authoritative data concerning blockchain uptake in health care. Hence, the snapshot provided by the study of current and ongoing industry engagement and experience with blockchain relies heavily on web-based investigations of gray literature, particularly industry publications, to investigate the focus and tenor of discussions about these technologies in health care and the trajectory of their development. This methodological approach was limited in several significant aspects, insofar as it relied on unconfirmed industry sources publicly available on English-language websites as a substantial source of evidence for understanding the current state of blockchain development and implementation in health care contexts. As the accuracy of claims made on public websites is practically impossible to verify and often fails to provide significant financial information about the level of investment or revenue generated (or expected to be generated), the findings on blockchain take-up in the health care sector are not comprehensive, particularly given the growing interest in developing blockchain applications in non–English-speaking countries (eg, China [171,172] and Japan [173]).

Although significant assurance about the reliability and robustness of the study's provisional findings was obtained by eliciting feedback from industry experts through the second focus group workshop while cross-checking the validity of these provisional findings against the published insights of industry specialists [174-176], the findings concerning the state of industry experimentation and maturity offered here are neither authoritative nor comprehensive. Nevertheless, they provide worthwhile insights into the general state of the health care sector's engagement with blockchain technologies to date, the particular domains and applications in which this engagement is occurring, and the current direction of travel. Accordingly, these findings offer a useful reference point for critical reflection in the context of future investigations of blockchain in health care and beyond.

Conclusions

This paper has demonstrated that the health care sector's engagement with blockchain technology has grown steadily since 2016, with the most promising activity occurring at the intra- and interorganizational level to enable data sharing and cooperation in health care administration and medical research. Although various blockchain initiatives have appeared in response to the COVID-19 pandemic [177], in which the rapid spread of misinformation intensified the importance of ensuring trustworthy health care information management, industry experts suggest that they have not hitherto taken root [178]. Contrary to the focus of the academic literature, the health care sector's engagement with blockchain has not focused primarily on enabling and integrating patient-controlled EHRs into clinical practice. It is in nonclinical contexts that early-stage health care engagement with blockchain is now producing evidence that blockchain-enabled cooperation can create both collective and individual value for network members that is greater than the sum of its parts, generating potentially significant efficiency gains [121]. Accordingly, blockchain in health care cannot currently be dismissed as a passing fad. Although scholars of innovation studies have identified various *barriers to innovation* that arise when technological inventions seek to cross the *valley of death* during the development phase of innovation, this literature has tended to treat technological inventions as a *black box*. In contrast, the analysis offered here highlights the importance of attending to the design and functionality of the technology itself and the particular health care contexts and

fields of practice into which it might be applied in ways that may deepen and enrich innovation studies literature that has tended toward a high level of generality and abstraction.

This study identified 6 complex and multidimensional challenges and 3 normative tensions that must be resolved if blockchain technologies are to be made readily implementable in health care. Negotiating and overcoming these challenges and tensions will require considerable commitment, including time, investment, and willingness by health care organizations to take substantial risks, experiment, learn, and share knowledge across the sector. It suggests that the technology's progress through the development process will be difficult and lengthy. Moreover, as the National Advisory Group on Health Information Technology advised in its 2016 report, "implementing health IT is one of the most complex adaptive changes in the history of healthcare, and perhaps of any industry" [179]. Another leading expert observes that "while policy makers are calling for technology to be implemented rapidly and at scale, the reality is that when dealing with the multiple complexities of health and care, it is extremely difficult to go beyond small-scale demonstration projects" [85].

Even if blockchain succeeds in crossing the valley of death, it does not necessarily follow that these technologies will live up to their promise. The underwhelming experience of health IT offers a sobering reminder, with Kellermann and Jones [180] noting that despite the increased use of health IT in the United States, with aggregate annual national expenditure on health care growing from US $2 trillion in 2005 to roughly US $2.8 trillion in 2013, the efficiency of patient care is only marginally better. In the United Kingdom, 10 years after the launch of the National Program for IT, which promised to revolutionize care in the English NHS [181,182], it was described by the UK Secretary of State for Health as a "huge disaster that became impossible to deliver" despite costing £3.66 billion (US $4.84 billion) [183,184]. These experiences highlight the serious gap between the grand visions of IT's transformative potential in health care and the fraught nature of implementation, given the inherently complex, dynamic, and context-driven nature of health care. They throw into high relief the inherent difficulties associated with seeking to *hard code* norms into complex sociotechnical systems in ways that that can satisfactorily resolve and mediate a wide variety of often conflicting rights, interests, expectations, norms, and risks associated with information sharing in dynamic, complex but safety-critical health care settings.

Although the *blockchain for health care promise* is rooted in a belief that blockchain systems will facilitate secure data sharing while enabling more fine-grained management of health care data, thereby creating *more and better data* that will drive improvements across the sector, real-world engagements and experiences with the technology in health care settings to date suggests that blockchain's core *value proposition* is more nuanced and narrowly framed. Blockchain's value generally (both within health care and beyond) rests primarily on its capacity to provide an automated, mathematically verifiable, highly secure, and therefore, highly trustworthy record-keeping function. It is this functional capacity that supports the technology's nascent potential as *middleware*, facilitating various kinds of health care data sharing and exchange, particularly between health care organizations where a lack of social trust rooted in institutional incentives for opportunistic behavior might be replaced with the computational trust that permissioned blockchain systems can provide. In so doing, blockchain systems may enable new forms of cooperation that generate shared value from the pooling of health care data, although whether these networks can develop governance strategies that can successfully and sustainably mediate the relevant interests of network participants remains to be seen.

However, the difficulties of ensuring that the ledger accurately, systematically, and comprehensively tracks and reflects the real-world phenomenon that it claims to record may mean that the envisaged benefits of blockchain's automated *track and trace* function may not be fully realized. Computer scientists and mathematicians have long recognized the *garbage in, garbage out* problem: the quality of output is determined by the quality of the input. Blockchain technology cannot guarantee that data arriving at and recorded in the ledger is an accurate, complete, and comprehensive record of the underlying real-world events that it purports to track; however, blockchain advocates typically overlook this critical vulnerability. In other words, although blockchain systems can deliver valuable functionality in terms of automatically and securely recording the handling of data items (for example, when a digital record has been viewed by a particular user), they cannot guarantee the integrity of the underlying data.

Nevertheless, blockchain's automated audit function might offer real and significant value because, by creating a highly secure and verifiable record, it offers a vehicle through which organizations can *demonstrate compliance with data handling norms* at scale, thus operating as a form of *RegTech*. In this way, blockchain as middleware can automatically and securely track and record the location and handling of data items. Therefore, it might have real potential to facilitate the *integration* of health data *silos* stored in a variety of locations across multiple organizations using different IT systems [185]. The value of this contribution could potentially be significant, reducing costs and improving efficiency, transparency, and security, and could be widely applied across any sector in which sensitive data generation, processing, and sharing occur in large volumes.

Blockchain is frequently portrayed as offering the potential to deliver radical and positive transformations to many sectors, including health care. This study has identified significant activity within the health care sector aimed at exploring how blockchain technologies might deliver real value. It has been found that the technology is still in its infancy and that the nature and scale of its value to health care remain unproven [29]. Experience of early health care sector engagement has shown that the relevant question is not *how can blockchain be used* in health care settings but *under what conditions can blockchain be realistically expected to provide a viable, scalable and efficient solution to a concrete health care problem in real-world health care settings that would add significant value relative to alternative solutions?* In other words, when considering whether to engage with blockchain technologies, organizations must ask themselves whether a blockchain is really necessary

to solve the problem at hand or whether other alternative approaches are preferable.

This shift in mindset is reflected in the divergence between the *rhetoric* of blockchain-enabled health care data sharing, which focuses primarily on the capacity of blockchain technologies to enable the seamless yet secure sharing of patient records that will enhance clinical decision-making, and the *reality of* engagement with blockchain, which has largely avoided clinical contexts. Instead, the most innovative health care blockchain applications have occurred at the interorganizational level, enabling the sharing of administrative data between health care organizations to improve the efficiency of health care administration. The most recent initiatives are aimed at facilitating shared access to the secondary use of patient data for medical research through cooperative endeavors to train algorithmic models on data sets in a secure, privacy-preserving manner.

However, to succeed in this endeavor, there remain very significant challenges associated with achieving social trust and cooperation, including the need for the effective and legitimate governance of blockchain networks, that no amount of technological sophistication or hard-coded solutions can, on their own, satisfactorily resolve. It remains to be seen whether these and other tensions inherent in these incipient forms of blockchain-enabled *coopetition* between health care organizations will be satisfactorily overcome and whether blockchain applications in health care might be characterized as *radical* or *incremental* technologies. Accordingly, the possibility of blockchain technology precipitating a *revolution* in the delivery of health care of the kind envisaged by blockchain for health care advocates appears to have been exaggerated. Nevertheless, there are real prospects of using blockchain in ways that could lead to genuine health care improvements. However, progress toward this goal is likely to occur slowly, incrementally, and more *by stealth* rather than through radical and rapid *disruption* over a short period and—at least in the first instance—likely to be concentrated across a much narrower range of applications than the blockchain for health care promise suggests.

Acknowledgments

The author gratefully acknowledges funding support from the Wellcome Trust, grant reference number: 210337/Z/18/Z, “Regulating healthcare through blockchain: Mapping the legal, ethical, technical and governance challenges.” The author is grateful for the research and administrative assistance provided by Immaculate Motsi-Omoijiade and Alexander Kharlamov and for the editing and proofreading assistance provided by PhD candidate Nathalie Smuha. The author is also indebted to those who participated in the focus group meetings convened for the purposes of this study, including John Halamka for his support and encouragement. Special thanks are due to colleagues who kindly read and commented on earlier drafts: Atina Krajewska, Muireann Quigley, Natalie Pankova, Barbara Prainsack, and Mark Taylor. All errors remain the author’s own. Further information about the project is available on the web [186].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of systematic literature reviews of blockchain in health care.
[DOCX File , 23 KB - jmir_v23i12e24109_app1.docx]

Multimedia Appendix 2
Summary of method and findings mapping blockchain in health care by Motsi-Omoijiade and Kharlamov [21].
[DOCX File , 53 KB - jmir_v23i12e24109_app2.docx]

Multimedia Appendix 3
Focus group participants and proceedings.
[DOCX File , 26 KB - jmir_v23i12e24109_app3.docx]

References

1. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
2. Khezr S, Moniruzzaman M, Yassine A, Benlamri R. Blockchain technology in healthcare: a comprehensive review and directions for future research. Appl Sci 2019 Apr 26;9(9):1736. [doi: 10.3390/app9091736]
3. Hölbl M, Kompara M, Kamišalić A, Zlatolas L. A systematic review of the use of blockchain in healthcare. Symmetry 2018 Oct 10;10(10):470. [doi: 10.3390/sym10100470]
4. Kuo T, Kim H, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
5. Blockchain for Healthcare Today. URL: https://blockchainhealthcaretoday.com [accessed 2021-01-12]

6. Zhang P, Schmidt D, White J, Lenz G. Blockchain technology use cases in healthcare. In: Advances in Computers. Amsterdam: Elsevier; 2018:1-41.
7. Durneva P, Cousins K, Chen M. The current state of research, challenges, and future research directions of blockchain technology in patient care: systematic review. J Med Internet Res 2020 Jul 20;22(7):e18619 [FREE Full text] [doi: 10.2196/18619] [Medline: 32706668]
8. Gaynor M, Tuttle-Newhall J, Parker J, Patel A, Tang C. Adoption of blockchain in health care. J Med Internet Res 2020 Sep 17;22(9):e17423 [FREE Full text] [doi: 10.2196/17423] [Medline: 32940618]
9. Mello MM, Adler-Milstein J, Ding KL, Savage L. Legal barriers to the growth of health information exchange-boulders or pebbles? Milbank Q 2018 Dec;96(1):110-143. [doi: 10.1111/1468-0009.12313] [Medline: 29504197]
10. Biot C, Johnson P, Massart S, Pecuchet N. Data sharing is the key to innovation in healthcare. MIT Technology Review. 2019. URL: https://www.technologyreview.com/s/614442/data-sharing-is-key-to-innovation-in-health-care/ [accessed 2021-12-01]
11. Orcutt M. Why the CDC wants in on blockchain. MIT Technology Review. 2017. URL: https://www.technologyreview.com/s/608959/why-the-cdc-wants-in-on-blockchain/ [accessed 2021-12-01]
12. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. JMIR Med Inform 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: 10.2196/11949] [Medline: 30578196]
13. Bass J. Introduction. In: Metcalf D, Bass J, Hooper M, Cahana A, Dhillon V, editors. Blockchain in Healthcare. Boca Raton, FL: CRC Press; 2019:3-24.
14. Halamka J, Lippman A, Ekblaw A. The potential for blockchain to transform electronic health records. Harvard Business Review. 2017. URL: https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records [accessed 2021-12-02]
15. Adler-Milstein J, Lin S, Jha A. The number of health information exchange efforts is declining, leaving the viability of broad clinical data exchange uncertain. Health Aff (Millwood) 2016 Jul 01;35(7):1278-1285 [FREE Full text] [doi: 10.1377/hlthaff.2015.1439] [Medline: 27385245]
16. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. Br Med J 2015 Mar 20;350(mar20 1):h1139 [FREE Full text] [doi: 10.1136/bmj.h1139] [Medline: 25794882]
17. Engelhardt M. Hitching healthcare to the chain: an introduction to blockchain technology in the healthcare sector. Technol Innovat Manag Rev (10) 2017;7 [FREE Full text] [doi: 10.22215/timreview/1111]
18. Agbo CC, Mahmoud QH, Eklund JM. Blockchain technology in healthcare: a systematic review. Healthcare (Basel) 2019 Apr 04;7(2) [FREE Full text] [doi: 10.3390/healthcare7020056] [Medline: 30987333]
19. Yaga D, Mell P, Roby N, Scarfone K. Blockchain technology overview. National Institute of Standards and Technology, Gaithersburg. 2018. URL: https://nvlpubs.nist.gov/nistpubs/ir/2018/nist.ir.8202.pdf [accessed 2021-12-01]
20. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
21. Motsi-Omoijiade I, Kharlamov A. Blockchain for healthcare applications and use-cases. In: De Gregorio G, Pollicino O, editors. Blockchain and Public Law. London, UK: Edward Elgar Publishing; 2021:136-189.
22. E-Estonia Heathcare. URL: https://e-estonia.com/solutions/healthcare/ [accessed 2021-12-02]
23. Miller R. An update on healthcare ICOs. Bertcmiller. 2019. URL: https://bertcmiller.com/2019/02/24/An-update-on-healthcare-ICOs.html [accessed 2021-12-02]
24. Healthcare and biomedical blockchains. Centre for Biomedical Blockchain Research. 2019. URL: https://db.biomedicalblockchain.org/companies_detail [accessed 2020-10-20]
25. Coravos A. Open-source landscape map for healthcare-related blockchains. GitHub. 2018. URL: https://github.com/acoravos/healthcare-blockchains [accessed 2018-05-30]
26. Stoffregen E. Blockchain Healthcare Ecosystem in 2018. Medium. 2018. URL: https://medium.com/@erikstoffregen/blockchain-healthcare-ecosystem-d21631024454 [accessed 2019-10-20]
27. Mire S. 40 Startups using blockchain to transform healthcare. Disruptor Daily. 2018. URL: https://www.disruptordaily.com/blockchain-market-map-healthcare/ [accessed 2019-10-20]
28. Listing of Healthcare ICOs (Initial Coin/Cryptoasset Offerings) / Token Offerings. Healthcare ICOs/Tokens-Medium. 2018. URL: https://docs.google.com/spreadsheets/d/1TANOZmuYtVhyn1C9PV6YLsOBlQNOEOBExB7u2_Kkork/edit#gid=0 [accessed 2021-12-08]
29. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?' - challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med 2019 Mar 27;17(1):68 [FREE Full text] [doi: 10.1186/s12916-019-1296-7] [Medline: 30914045]
30. Metcalf D, Bass J, Hooper M, Cahana A, Dhillon V. Blockchain in Healthcare: Innovations That Empower Patients, Connect Professionals and Improve Care. Boca Raton, FL: CRC Press; 2019.
31. Vigna P, Casey M. The Truth Machine: The Blockchain and the Future of Everything. London: Picador; 2019.
32. El-Gazzar R, Stendal K. Blockchain in health care: hope or hype? J Med Internet Res 2020 Jul 10;22(7):e17199 [FREE Full text] [doi: 10.2196/17199] [Medline: 32673219]

33. Abernathy W, Utterback J. Patterns of industrial innovation. Technol Rev 1978;80(7):40-47 [FREE Full text]
34. Rosenberg N, Nathan R. Inside the Black Box: Technology and Economics. Cambridge: Cambridge University Press; 1982.
35. Sahal D. Technological guideposts and innovation avenues. Res Policy 1985 Apr;14(2):61-82. [doi: 10.1016/0048-7333(85)90015-0]
36. Bums T, Stalker G. The Management of Innovation. Oxfordshire, United Kingdom: Oxford Scholarship Online; 1994.
37. Abernathy W. The Productivity Dilemma: Roadblock to Innovation in the Automobile Industry. Baltimore, MD: Johns Hopkins University Press; 1978.
38. Biemans W, Huizingh K. Mapping the valley of death ecosystem. In: Bitran I, Conn S, Gernreich C, Heber M, Huizingh E, Kokshagina O, et al, editors. Proceedings of ISPIM Connects Bangkok: Partnering for an Innovative Community. Bangkok: ISPIM; 2020.
39. Markham S. Moving technologies from lab to market. Res Technol Manag 2016 Jan 27;45(6):31-42 [FREE Full text] [doi: 10.1080/08956308.2002.11671531]
40. Helmstetter M. Bridging the valley of death between Ag Innovation Funding and market adoption. Forbes. 2018. URL: https://www.forbes.com/sites/michaelhelmstetter/2018/08/03/bridging-the-valley-of-death-between-innovation-funding-and-market-adoption/ [accessed 2021-12-02]
41. Broens T, Huis in't Veld RM, Vollenbroek-Hutten MM, Hermens HJ, van Halteren AT, Nieuwenhuis LJ. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007;13(6):303-309 [FREE Full text] [doi: 10.1258/135763307781644951] [Medline: 17785027]
42. Parrish M, Tan Y, Grimes K, Mochly-Rosen D. Surviving in the valley of death: opportunities and challenges in translating academic drug discoveries. Annu Rev Pharmacol Toxicol 2019 Jan 06;59:405-421 [FREE Full text] [doi: 10.1146/annurev-pharmtox-010818-021625] [Medline: 30208282]
43. Branscomb L, Morse K, Robert M. Managing technical risk: Understanding private sector decision making on early stage technology-based projects. Prepared for U.S. Department of Commerce Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology, NIST GCR 00-787. 2000. URL: http://www.atp.nist.gov/eao/gcr_787.pdf [accessed 2021-12-02]
44. Sandberg B, Aarikka-Stenroos L. What makes it so difficult? A systematic review on barriers to radical innovation. Ind Mark Manag 2014 Nov;43(8):1293-1305 [FREE Full text] [doi: 10.1016/j.indmarman.2014.08.003]
45. Grisot M, Hanseth O, Thorseng A. Innovation of, in, on infrastructures: articulating the role of architecture in information infrastructure evolution. J Assoc Info Syst 2014 Apr;15(4):197-219. [doi: 10.17705/1jais.00357]
46. Slayton R, Spinardi G. Radical innovation in scaling up: Boeing's Dreamliner and the challenge of socio-technical transitions. Technovation 2016 Jan;47:47-58 [FREE Full text] [doi: 10.1016/j.technovation.2015.08.004]
47. Berg M, Timmermans S. Orders and their others: on the constitution of universalities in medical work. Configurations 2000;8(1):31-61. [doi: 10.1353/con.2000.0001]
48. Maniatopoulos G, Procter R, Llewellyn S, Harvey G, Boyd A. Moving beyond local practice: reconfiguring the adoption of a breast cancer diagnostic technology. Soc Sci Med 2015 Apr;131:98-106 [FREE Full text] [doi: 10.1016/j.socscimed.2015.02.036] [Medline: 25769108]
49. Berg M. On distribution, drift and the electronic medical record. In: Proceedings of the Fifth European Conference on Computer Supported Cooperative Work. 1997 Presented at: Fifth European Conference on Computer Supported Cooperative Work; Sept 7-11, 1997; Lancaster, UK p. 141-156. [doi: 10.1007/978-94-015-7372-6_10]
50. Greenhalgh T, Potts H, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. Milbank Q 2009 Dec;87(4):729-788 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00578.x] [Medline: 20021585]
51. EU-funded projects in blockchain technology. European Commission. 2019. URL: https://digital-strategy.ec.europa.eu/en/news/eu-funded-projects-blockchain-technology [accessed 2021-12-03]
52. Shen M. UK Government-backed accelerator will sponsor blockchain startups. Coindesk. 2018. URL: https://www.coindesk.com/uk-government-backed-accelerator-will-sponsor-blockchain-startups; [accessed 2021-12-02]
53. Miller R, Kuraitis V. The biggest trend you've probably never heard of: a status report on 138 healthcare ICOs. The Health Care Blog. 2018. URL: https://thehealthcareblog.com/blog/2018/09/24/the-biggest-trend-youve-probably-never-heard-of-a-status-report-on-138-healthcare-icos/ [accessed 2021-12-02]
54. John J. State of blockchain Q3: the professionals have moved in with VC investments soaring to an all time high. Outlier Ventures. 2018. URL: https://outlierventures.io/research/state-of-blockchains-q3-the-professionals-have-moved-in-with-vc-investments-soaring-to-all-time-high/ [accessed 2021-12-02]
55. Miller R. Blockchain in healthcare in 2019. Bertcmiller. 2019. URL: https://www.bertcmiller.com/2019/12/29/blockchain-healthcare-2019.html [accessed 2021-12-02]
56. Clauson K, Breeden E, Davidson C, Mackey T. Leveraging Blockchain Technology to Enhance Supply Chain Management in Healthcare: an exploration of challenges and opportunities in the health supply chain. Blockchain Healthc Today 2018 Mar 23;1 [FREE Full text] [doi: 10.30953/bhty.v1.20]
57. Bass J. The seven major consortia (in chronological order). Hashed Health Newsletter. 2019. URL: https://hashedhealth.com/consortia-july-2019-2/ [accessed 2021-12-02]

58. Bass J. The 2018 stories that shaped blockchain in healthcare. Hashed Health Newsletter. 2019. URL: https://hashedhealth.com/2018_hhighlights/ [accessed 2021-12-02]
59. DCSMA Pilot Project. US Food and Drug Administration. 2019. URL: https://www.fda.gov/drugs/drug-supply-chain-security-act-dscsa/dscsa-pilot-project-program [accessed 2021-12-02]
60. Chien W, De Jesus J, Taylor B, Dods V, Alekseyev L, Shoda D, et al. The last mile: DSCSA solution through blockchain technology: drug tracking, tracing, and verification at the last mile of the pharmaceutical supply chain with BRUINchain. Blockchain Healthc Today 2020 Mar 12 [FREE Full text] [doi: 10.30953/bhty.v3.134]
61. Woods J, Iyengar-Emens R. Blockchain to secure a more personalized pharma. Genet Eng Biotechnol News 2019 Jan;39(1):27-29 [FREE Full text] [doi: 10.1089/gen.39.01.08]
62. Subramanian N, Chaudhuri A, Kayikci Y. Blockchain applications in health care supply chain. In: Blockchain and Supply Chain Logistics. Cham: Palgrave Pivot; 2020:31-38.
63. Walport M. FinTech futures: The UK as a world leader in financial technologies. Report to UK Government Office for Science. 2015. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/413095/gs-15-3-fintech-futures.pdf [accessed 2021-12-02]
64. Yeung K. Regulation by blockchain: the emerging battle for supremacy between the code Law and Code Law. Modern Law Rev 2019 Mar 05;82(2):207-239 [FREE Full text] [doi: 10.1111/1468-2230.12399]
65. Fassenbender M. Pfizer, Biogen-led blockchain project demonstrates proof of concept in pursuit of a better supply chain. Outsourcing-Pharma. 2019. URL: https://www.outsourcing-pharma.com/Article/2019/08/14/Clinical-Supply-Blockchain-Working-Group-led-by-Pfizer-and-Biogen-completes-proof-of-concept-project [accessed 2021-12-02]
66. Henderson L. Blockchain in research, from business to personal. Applied Clinical Trials (11). 2018. URL: https://search.proquest.com/scholarly-journals/blockchain-research-business-personal/docview/2153565380/se-2?accountid=8630 [accessed 2021-12-03]
67. Synaptic Health Alliance. URL: https://www.synaptichealthalliance.com/project [accessed 2021-12-03]
68. Begando A. Developing a professional credentials exchange levering distributed ledger technology? In: Metcalf D, Bass J, Hooper M, Cahana A, Dhillon V, editors. Blockchain in Healthcare. Boca Raton, FL: CRC Press; 2019:3-24.
69. See CJ. How ProCredEx is leveraging blockchain to speed up credentialing. Mobile Health News. 2018. URL: https://www.mobihealthnews.com/content/how-procredex-leveraging-blockchain-speed-credentialing [accessed 2021-12-03]
70. Armstrong S. Bitcoin technology could take a bite out of NHS data problem. Br Med J 2018 May 08:k1996 [FREE Full text] [doi: 10.1136/bmj.k1996]
71. National Institute for Health and Care Excellence (NICE). URL: https://www.nice.org.uk/ [accessed 2021-12-03]
72. The NHS Long Term Plan. National Health Service. 2019. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2021-12-03]
73. Bourne J. Six Questions with Sam Shah, Director of Digital Development, NHS England. Blockchain Expo World Series. 2019. URL: https://www.blockchain-expo.com/2019/02/blockchain/blockchain-expo-global-nhs/ [accessed 2021-12-03]
74. Dovetail Labs. URL: https://dovetaillabs.com/ [accessed 2021-12-03]
75. Acquisition of Dovetail Digital Limited. EMIS Group. 2018. URL: https://www.emisgroupplc.com/news-and-media/news/acquisition-of-dovetail-digital-limited/ [accessed 2021-12-03]
76. World's first blockchain-supported Personal Care Record Platform launched by Guardtime and partners to up to 30 million NHS patients in the UK. Guardtime. 2018. URL: https://guardtime.com/blog/world-s-first-blockchain-supported-personal-care-record-platform-launched-by-guardtime-and-partners [accessed 2021-12-03]
77. Medicalchain announces partnership with the Grove Medical Group. Journal of mHealth. 2018. URL: https://thejournalofmhealth.com/medicalchain-announces-partnership-groves-medical-group/ [accessed 2021-12-03]
78. Medicalchain and Better by Marrand announce partnership. Medicalchain. 2019. URL: https://medicalchain.com/en/medicalchain-and-better-by-marand-announce-partnership/ [accessed 2021-12-03]
79. Passerat-Palmbach J, Farnan T, Miller R, Gross MS, Flannery HL, Gleim B. A blockchain-orchestrated Federated Learning architecture for healthcare consortia. arXiv. 2019. URL: https://arxiv.org/abs/1910.12603 [accessed 2021-12-03]
80. Miller R. Why federated learning is important. 2020. URL: https://bertcmiller.com/2020/05/17/why-federated-learning-is-important.html [accessed 2021-12-08]
81. Greenhalgh T, Swinglehurst D, Stones R. Rethinking resistance to 'big IT': a sociological study of why and when healthcare staff do not use nationally mandated information and communication technologies. In: Health Services and Delivery Research. Southampton (UK): NIHR Journals Library; 2014.
82. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: systematic review. J Med Internet Res 2019 May 10;21(5):e12426 [FREE Full text] [doi: 10.2196/12426] [Medline: 31094344]
83. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: a systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250 [FREE Full text] [doi: 10.1258/jhsrp.2010.009137] [Medline: 20592046]

84. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Dec 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
85. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
86. Brynjolfsson E. The productivity paradox of information technology. Commun ACM 1993 Dec;36(12):66-77. [doi: 10.1145/163298.163309]
87. Greenhalgh T, Fahy N, Shaw S. The bright elusive butterfly of value in health technology development comment on "Providing value to new health technology: the early contribution of entrepreneurs, investors, and regulatory agencies". Int J Health Policy Manag 2018 Jan 01;7(1):81-85 [FREE Full text] [doi: 10.15171/ijhpm.2017.65] [Medline: 29325407]
88. Chang SE, Chen Y. Blockchain in health care innovation: literature review and case study from a business ecosystem perspective. J Med Internet Res 2020 Aug 31;22(8):e19480 [FREE Full text] [doi: 10.2196/19480] [Medline: 32865501]
89. Hau YS, Lee JM, Park J, Chang MC. Attitudes toward blockchain technology in managing medical information: survey study. J Med Internet Res 2019 Dec 09;21(12):e15870 [FREE Full text] [doi: 10.2196/15870] [Medline: 31815676]
90. Lee K, Lim K, Jung SY, Ji H, Hong K, Hwang H, et al. Perspectives of patients, health care professionals, and developers toward blockchain-based health information exchange: qualitative study. J Med Internet Res 2020 Nov 13;22(11):e18582 [FREE Full text] [doi: 10.2196/18582] [Medline: 33185553]
91. Bass J. Healthcare blockchain dreams: 5 things people thought would happen by 2020. Hashed Health. 2019. URL: https://hashedhealth.com/5-things-by-2020/ [accessed 2021-12-03]
92. Considerations for Policymakers: The Application of Blockchain to Healthcare. HIMSS. 2019. URL: https://www.himss.org/resources/considerations-policymakers-application-blockchain-technology-healthcare [accessed 2021-12-03]
93. Finck M. Blockchain Regulation and Governance in Europe. Cambridge University Press: Cambridge; 2018.
94. Werbach K. Trustless trust. SSRN J 2016:31 [FREE Full text] [doi: 10.2139/ssrn.2844409]
95. Szabo N. Formalizing and securing relationships on public networks. First Monday 1997;2(9) [FREE Full text] [doi: 10.5210/fm.v2i9.548]
96. Arner D, Barberis J, Buckey R. FinTech, RegTech, and the reconceptualization of financial regulation. Nw J Int'l L & Bus 2017;37:371 [FREE Full text]
97. Walport M. Distributed ledger technology: beyond block chain. UK Government Office for Science. 2016. URL: https://www.gov.uk/government/news/distributed-ledger-technology-beyond-block-chain [accessed 2021-12-03]
98. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019 Dec 19;21(12):e16093 [FREE Full text] [doi: 10.2196/16093] [Medline: 31855184]
99. Shull JG. Digital health and the state of interoperable electronic health records. JMIR Med Inform 2019 Nov 01;7(4):e12712 [FREE Full text] [doi: 10.2196/12712] [Medline: 31682583]
100. Interoperability and blockchain. HIMSS (No longer open access on 15.1.21). URL: https://www.himss.org/interoperability-and-blockchain-healthcare [accessed 2021-12-03]
101. Oosterwijk H. Determining the measures of success for interoperability. HIMSS. 2019. URL: https://www.himss.org/resources/determining-measures-success-interoperability [accessed 2021-12-03]
102. Interoperability in healthcare. HIMSS. URL: https://www.himss.org/resources/interoperability-healthcare [accessed 2021-12-04]
103. IHE International. IHE USA. URL: https://www.iheusa.org/ihe-international [accessed 2021-12-04]
104. HL7. Health Level Seven International. URL: http://www.hl7.org/about/index.cfm?ref=nav [accessed 2021-12-04]
105. Zhang PW, White J, Schmidt DC, Lenz G, Rosenbloom ST. FHIRChain: applying blockchain to securely and scalably share clinical data. Comput Struct Biotechnol J 2018;16:267-278 [FREE Full text] [doi: 10.1016/j.csbj.2018.07.004] [Medline: 30108685]
106. Beyond Blocks biennial blockchain in healthcare report. Miller R. 2019. URL: https://bertcmiller.com/2019/07/07/Beyond-Blocks-Report.html [accessed 2021-12-03]
107. Bass J. Healthcare blockchain consortia initiatives. Hashed Health Newsletter. URL: https://hashedhealth.com/newsletter-july-2019/ [accessed 2021-12-03]
108. Miller R. A note on the challenges in interoperability between pharmaceutical supply chain blockchain networks. 10 May 2020. URL: https://bertcmiller.com/2020/05/10/interoperability-supply-chain-blockchain.html [accessed 2021-12-03]
109. Dubovitskaya A, Baig F, Xu Z, Shukla R, Zambani PS, Swaminathan A, et al. ACTION-EHR: patient-centric blockchain-based electronic health record data management for cancer care. J Med Internet Res 2020 Aug 21;22(8):e13598 [FREE Full text] [doi: 10.2196/13598] [Medline: 32821064]
110. Miller R. Top posts of 2019. Bertcmiller. 2020. URL: https://www.bertcmiller.com/2019/12/29/blockchain-healthcare-2019.html [accessed 2021-12-03]

111. Walch A. Open-source operational risk: should public blockchains serve as financial market infrastructures? In: Chen D, Deng RH, editors. Handbook of Blockchain, Digital Finance, and Inclusion. London: Academic Press; 2018:243-269.
112. Vidan G, Lehdonvirta V. Mine the gap: Bitcoin and the maintenance of trustlessness. New Media Society 2018 Jul 11;21(1):42-59 [FREE Full text] [doi: 10.1177/1461444818786220]
113. Dodd N. The social life of Bitcoin. Theory Cult Society 2017 Dec 17;35(3):35-56 [FREE Full text] [doi: 10.1177/0263276417746464]
114. De Fillippi P, Loveluck B. The invisible politics of bitcoin: governance crisis of a decentralized infrastructure. Internet Policy Rev 2016;5(4):35-56. [doi: 10.14763/2016.3.427]
115. Yeung K, Galindo D. Why do public blockchains need formal and effective internal governance mechanisms? Eur J Risk Regul 2019 Sep 18;10(2):359-375 [FREE Full text] [doi: 10.1017/err.2019.42]
116. Howell BE, Potgieter PH, Sadowski BM. Governance of blockchain and distributed ledger technology projects. SSRN J 2020. [doi: 10.2139/ssrn.3365519]
117. Rosenbaum L. Anthem will use blockchain to secure medical data for its 40 million members in three years. Forbes. 2019 Dec 12. URL: https://tinyurl.com/2p9dmxxj [accessed 2021-12-03]
118. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. J Med Internet Res 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: 10.2196/12533] [Medline: 30735142]
119. Wüst K, Gervais A. Do you need a blockchain? In: Proceedings of the Crypto Valley Conference on Blockchain Technology (CVCBT).: IEE; 2018 Nov 08 Presented at: Crypto Valley Conference on Blockchain Technology (CVCBT); 2018; Zug, Switzerland p. 45-54. [doi: 10.1109/CVCBT.2018.00011]
120. Garrard R, Fielke S. Blockchain for trustworthy provenances: a case study in the Australian aquaculture industry. Technol Soc 2020 Aug;62:101298 [FREE Full text] [doi: 10.1016/j.techsoc.2020.101298]
121. Houlding D. Thought leader perspectives: an interview with David Houlding on blockchain, enterpriseHealth. In: Metcalf D, Bass J, Hooper M, Cahana A, Dhillon V, editors. Blockchain in Healthcare. Boca Raton, FL: CRC Press; 2019:167-177.
122. Wachter R. The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. New York: McGraw-Hill Education; 2015.
123. Flannery H. Standards. In: Metcalf D, Bass J, Hooper M, Cahana A, Dhillon V, editors. Blockchain in Healthcare. Boca Raton, FL: CRC Press; 2019:233.
124. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019 May 13;2(1):38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
125. NHS Clinical Digital Safety Team, Standard DCB0129: Clinical Risk Management: its application in the manufacture of health IT systems. NHS Digital. 2020. URL: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0129-clinical-risk-management-its-application-in-the-manufacture-of-health-it-systems [accessed 2021-12-03]
126. NHS Clinical Digital Safety Team. Clinical Risk Management: Its Application in the Deployment and Use of Health IT Systems. NHS Digital. 2020. URL: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0160-clinical-risk-management-its-application-in-the-deployment-and-use-of-health-it-systems [accessed 2021-12-03]
127. Qin K, Gervais A. An Overview of Blockchain Scalability, Interoperability and Sustainability. Hochschule Luzern, Imperial College London: Liquidity Network; 2018.
128. Blockchain in healthcare. HIMMS. URL: https://www.himss.org/resources/blockchain-healthcare [accessed 2021-12-03]
129. Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustainable Cities Society 2018 May;39:283-297. [doi: 10.1016/j.scs.2018.02.014]
130. Sater S. Blockchain transforming healthcare data flows. SSRN Network. 2020. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3171005 [accessed 2021-12-03]
131. Kyratsis Y, Ahmad R, Holmes A. Technology adoption and implementation in organisations: comparative case studies of 12 English NHS Trusts. BMJ Open 2012 Apr 04;2(2):e000872 [FREE Full text] [doi: 10.1136/bmjopen-2012-000872] [Medline: 22492183]
132. Kelly J. Blockchain insiders tell us why we don't need blockchain. Financial Times. 2018 Jun 14. URL: https://www.ft.com/content/469d9dae-3d82-344d-a8f2-e5376dbf4517 [accessed 2021-12-03]
133. Lehdonvirta V. The blockchain paradox: why distributed ledger technologies may do little to transform the economy. OII Blog. 2016 Nov 21. URL: https://www.oii.ox.ac.uk/news-events/news/the-blockchain-paradox-why-distributed-ledger-technologies-may-do-little-to-transform-the-economy/ [accessed 2021-12-03]
134. Bengtsson M, Raza-Ullah T. A systematic review of research on coopetition: toward a multilevel understanding. Ind Mark Manag 2016 Aug;57:23-39 [FREE Full text] [doi: 10.1016/j.indmarman.2016.05.003]
135. Brandenburger A, Nalebuff BJ. Co-opetition. London: HarperCollins Business; 1996.
136. Le Roy F, Czakon W. Managing coopetition: the missing link between strategy and performance. Ind Mark Manag 2016 Feb;53:3-6. [doi: 10.1016/j.indmarman.2015.11.005]

137. Le Roy F, Fernandez A, Chiambaretto P. From strategizing coopetition to managing coopetition. In: Fernandez AS, Chiambaretto P, Le Roy F, Czakon W, editors. The Routledge Companion to Coopetition Strategies. Abingdon: Taylor & Francis; 2018:36-46.
138. Fernandez A, Le Roy F, Gnyawali D. Sources and management of tension in co-opetition case evidence from telecommunications satellites manufacturing in Europe. Ind Mark Manag 2014 Feb;43(2):222-235 [FREE Full text] [doi: 10.1016/j.indmarman.2013.11.004]
139. Czakon W, Srivastava MK, Le Roy F, Gnyawali D. Coopetition strategies: critical issues and research directions. Long Range Plan 2020 Feb;53(1):101948. [doi: 10.1016/j.lrp.2019.101948]
140. Yami S, Castaldo S, Dagnino B, Le Roy F. Coopetition: Winning Strategies for the 21st Century. Cheltenham: Edward Elgar Publishing; 2010.
141. Avdoshin S, Pesotskaya E. Blockchain revolution in the healthcare industry. In: Proceedings of the Future Technologies Conference.: Springer; 2018 Presented at: Future Technologies Conference; November 15-16, 2018; Vancouver p. 626-639. [doi: 10.1007/978-3-030-02686-8_47]
142. Gordon W, Catalini C. Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Computat Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
143. Schneier B. There's no good reason to trust blockchain technology. Wired. 2019 Jun 02. URL: https://www.wired.com/story/theres-no-good-reason-to-trust-blockchain-technology/amp [accessed 2021-12-03]
144. Freeman C, Soete L. The Economics of Industrial Innovation. London: Frances Pinter; 1997.
145. Nelson R, Winter S. The Schumpeterian trade-off revisited. Am Econ Rev 1982;72(1):114-132.
146. Ettlie J, Bridges W, O'Keefe R. Organization strategy and structural differences for radical versus incremental i. Manag Sci 1984 Jun;30(6):682-695 [FREE Full text] [doi: 10.1287/mnsc.30.6.682]
147. Dewar R, Dutton J. The adoption of radical and incremental innovations: an empirical analysis. Manag Sci 1986 Nov;32(11):1422-1433 [FREE Full text] [doi: 10.1287/mnsc.32.11.1422]
148. Tushman M, Anderson P. Technological discontinuities and organizational environments. Admin Sci Q 1986 Sep;31(3):439-465 [FREE Full text] [doi: 10.2307/2392832]
149. Dess GG, Beard DW. Dimensions of organizational task environments. Admin Sci Q 1984 Mar;29(1):52. [doi: 10.2307/2393080]
150. Hage J. Theories of Organizations: Form, Process, and Transformation. Hoboken, New Jersey, United States: John Wiley & Sons; 1980.
151. Cooper A, Schendel D. Strategic responses to technological threats. Bus Horiz 1976 Feb;19(1):61-69 [FREE Full text] [doi: 10.1016/0007-6813(76)90024-0]
152. Daft R. Bureaucratic versus nonbureaucratic structure and the process of innovation and change. Res Sociol Organ 1982;1(1):129-166. [doi: 10.1108/s0733-558x20140000042019]
153. Rothwell R. Innovation and re-innovation: a role for the user. J Mark Manag 1986 Jan;2(2):109-123 [FREE Full text] [doi: 10.1080/0267257x.1986.9964004]
154. Christensen CM. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Boston: Harvard Business Review Press; 2015.
155. Henderson RM, Clark KB. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. Admin Sci Q 1990 Mar;35(1):9-30 [FREE Full text] [doi: 10.2307/2393549]
156. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Mar;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
157. Berg M. The search for synergy: interrelating medical work and patient care information systems. Methods Inf Med 2018 Feb 08;42(04):337-344. [doi: 10.1055/s-0038-1634227]
158. Maniatopoulos G, Llewellyn S, Procter R, Harvey G. Lost in translation?: negotiating technological innovation in healthcare. In: Proceedings of the European Group for Organisational Studies Colloquium. 2011 Presented at: The European Group for Organisational Studies Colloquium; 2011; Gotenberg p. 1-20.
159. Smaglik P. Fixing a broken record. Nature 2005 Jun 23;435(7045):1131. [doi: 10.1038/nj7045-1131a] [Medline: 15981321]
160. Considerations for policymakers: the application of blockchain in healthcare. HIMMS. 2019 Dec. URL: https://www.himss.org/resources/considerations-policymakers-application-blockchain-technology-healthcare [accessed 2021-12-03]
161. Wajcman J. Pressed for Time: The Acceleration of Life in Digital Capitalism. London: University of Chicago Press; 2015.
162. Alexaki S, Alexandris G, Katos V, Petroulakis NE. Blockchain-based electronic patient records for regulated circular healthcare jurisdictions. In: Proceedings of the IEEE 23rd International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD).: IEEE; 2018 Presented at: IEEE 23rd International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD); 2018; Barcelona, Spain p. 1-6. [doi: 10.1109/camad.2018.8514954]
163. German Federal Ministry of the Interior. Building and Community, Federal Ministry of Justice and Consumer Protection. Opinion of the Data Ethics Commission. 2019. URL: https://www.bmjv.de/DE/Themen/FokusThemen/Datenethikkommission/Datenethikkommission_EN_node.html;jsessionid=0B3B04760010083EFF159EE67CA14572.1_cid289 [accessed 2021-12-02]

164. Lu C, Batista D, Hamouda H, Lemieux V. Consumers' intentions to adopt blockchain-based personal health records and data sharing: Focus Group Study. JMIR Form Res 2020 Nov 05;4(11):e21995 [FREE Full text] [doi: 10.2196/21995] [Medline: 33151149]
165. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]
166. Morozov E. To Save Everything, Click Here: The Folly of Technological Solutionism. London: Allen Lane; 2013.
167. Lynskey O. The Foundations of EU Data Protection Law. Oxford: Oxford University Press; 2015:177-253.
168. Lanier J. Who Owns the Future?. New York: Simon and Schuster; 2014.
169. Schwartz PM. Property, privacy, and personal data. Harvard Law Rev 2004 May;117(7):2056. [doi: 10.2307/4093335]
170. Lin SC, Lyles C, Sarkar U, Adler-Milstein J. Are patients electronically accessing their medical records? Evidence from national hospital data. Health Aff (Millwood) 2019 Nov;38(11):1850-1857 [FREE Full text] [doi: 10.1377/hlthaff.2018.05437] [Medline: 31682494]
171. Jin X, Zhang M, Zhou Z, Yu X. Application of a blockchain platform to manage and secure personal genomic data: a case study of LifeCODE.ai in China. J Med Internet Res 2019 Sep 10;21(9):e13587 [FREE Full text] [doi: 10.2196/13587] [Medline: 31507268]
172. Wang Q, Su M, Li R. Is China the world’s blockchain leader? Evidence, evolution and outlook of China’s blockchain research. J Clean Prod 2020 Aug;264:121742 [FREE Full text] [doi: 10.1016/j.jclepro.2020.121742]
173. Mackey T, Bekki H, Matsuzaki T, Mizushima H. Examining the Potential of Blockchain Technology to Meet the Needs of 21st-Century Japanese Health Care: Viewpoint on Use Cases and Policy. J Med Internet Res 2020 Jan 09;22(1):e13649 [FREE Full text] [doi: 10.2196/13649] [Medline: 31917371]
174. Hashed Health. URL: https://hashedhealth.com/about/ [accessed 2021-12-01]
175. Robert Miller. URL: https://bertcmiller.com/ [accessed 2021-12-02]
176. Blockchain in healthcare and the life sciences. Consensys. URL: https://consensys.net/blockchain-use-cases/healthcare-and-the-life-sciences/ [accessed 2021-12-03]
177. Khurshid A. Applying blockchain technology to address the crisis of trust during the COVID-19 pandemic. JMIR Med Inform 2020 Sep 22;8(9):e20477 [FREE Full text] [doi: 10.2196/20477] [Medline: 32903197]
178. Miller R. Top 5 blockchain and healthcare trends in 2020: on business networks, the effect of COVID, and the direction of the market. Bertcmiller. URL: https://tinyurl.com/yckt223b [accessed 2021-12-03]
179. Wachter R. National Advisory Group on health information technology in England. In: Making It Work: Harnessing the Power of Health Information Technology to Improve Care in England. London: Department of Health. London: Department of Health; 2016.
180. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Aff (Millwood) 2013 Jan;32(1):63-68 [FREE Full text] [doi: 10.1377/hlthaff.2012.0693] [Medline: 23297272]
181. Department of Health. Liberating the NHS: An Information Revolution. London: HMSO; 2010.
182. Neville S. Jeremy Hunt reveals push for more IT use in NHS. Financial Times. 2013. URL: https://www.ft.com/content/60861f74-c3b5-11e2-8c30-00144feab7de [accessed 2021-12-03]
183. Greenhalgh T, Keen J. England's national programme for IT. Br Med J 2013 Jun 28;346(jun28 2):f4130. [doi: 10.1136/bmj.f4130] [Medline: 23810959]
184. Campion-Awwad O, Hayton A, Smith L, Vuaran M. The National Programme for IT in the NHS. A Case History, University of Cambridge. 2014. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.589.4954&rep=rep1&type=pdf [accessed 2021-12-02]
185. Healthcare Information and Management Systems Society (HIMSS). URL: https://www.himss.org/blockchain-deployment-architecture [accessed 2021-12-02]
186. Blockchain for Healthcare. URL: https://kyeung8.wixsite.com/blockchain4health [accessed 2021-12-08]

Abbreviations

B2B: business-to-business
EHR: electronic health record
HIMSS: Healthcare Information Management and Systems Society
IT: information technology
NHS: National Health Service
ProCredEx: Professionals Credentials Exchange
TAM: technology acceptance model

Edited by G Eysenbach; submitted 04.09.20; peer-reviewed by L Martin, F Prasser, I Mircheva; comments to author 14.11.20; revised version received 16.03.21; accepted 03.04.21; published 20.12.21.

Please cite as:
Yeung K
The Health Care Sector's Experience of Blockchain: A Cross-disciplinary Investigation of Its Real Transformative Potential
J Med Internet Res 2021;23(12):e24109
URL: https://www.jmir.org/2021/12/e24109
doi:10.2196/24109
PMID:34932009

Original Paper

Citation Advantage of Promoted Articles in a Cross-Publisher Distribution Platform: 36-Month Follow-up to a Randomized Controlled Trial

Paul Kudlow[1*], MD, PhD; Tashauna Brown[2*], MSc, PhD; Gunther Eysenbach[3*], MD, MPH

[1]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[2]TrendMD Inc, Toronto, ON, Canada
[3]University of Victoria, Victoria, BC, Canada
[*]all authors contributed equally

Corresponding Author:
Paul Kudlow, MD, PhD
Department of Psychiatry
University of Toronto
250 College Street
Toronto, ON, M5T 1R8
Canada
Phone: 1 416 480 6100
Email: paul.kudlow@gmail.com

Abstract

Background: There are limited evidence-based strategies that have been shown to increase the rate at which peer-reviewed articles are cited. In a previously reported randomized controlled trial, we demonstrated that promotion of article links in an online cross-publisher distribution platform (TrendMD) persistently augments citation rates after 12 months, leading to a statistically significant 50% increase in citations relative to the control.

Objective: This study aims to investigate if the citation advantage of promoted articles upholds after 36 months.

Methods: A total of 3200 published articles in 64 peer-reviewed journals across 8 subject areas were block randomized at the subject level to either the TrendMD group (n=1600) or the control group (n=1600) of the study. Articles were promoted in the TrendMD Network for 6 months. We compared the citation rates in both groups after 36 months.

Results: At 36 months, we found the citation advantage endured; articles randomized to TrendMD showed a 28% increase in mean citations relative to the control. The difference in mean citations at 36 months for articles randomized to TrendMD versus the control was 10.52 (95% CI 3.79-17.25) and was statistically significant (*P*=.001).

Conclusions: To our knowledge, this is the first randomized controlled trial to demonstrate how a postpublication article promotion intervention can be used to persistently augment citations of peer-reviewed articles. TrendMD is an efficient digital tool for knowledge translation and dissemination to targeted audiences to facilitate the uptake of research.

(J Med Internet Res 2021;23(12):e34051) doi:10.2196/34051

KEYWORDS

knowledge translation; knowledge; dissemination; digital knowledge translation; digital publishing; e-publishing; open access; scientometrics; infometrics

Introduction

Citations are a leading indicator of scholarly impact. They measure the spread of new knowledge; acknowledge the contribution of colleagues; and, in many fields, form the basis of tenure and promotion or even direct compensation [1,2]. Citations accrue from research getting noticed and used by authors when creating their own scholarly work. The problem is, there is a growing mismatch between the number of newly published articles and the number of papers an academic can discover and ultimately cite in the creation of their own scholarly work. There are over 8000 new peer-reviewed articles published daily, a figure that is growing exponentially, making it challenging for scholars to sift through potentially relevant

literature [3,4]. As a result, many relevant papers that could be cited are missed by scholars [3]. Accordingly, data suggests that roughly 35% of articles published between 1990 and 2015 remain uncited; a figure that may be increasing [5]. A study published in 2021 found that the increasing enormity of academic literature may be impeding on the rise, progress, and adoption of new ideas; the authors suggest that changes are needed in scholarship dissemination to increase the discoverability of new concepts that researchers may not be searching for [6]. The growing mismatch between the number of articles published and the ability for readers to come across articles that they were not already seeking out has created a pressing need for strategies that augment the discoverability of scholarly content and to target content to potential knowledge users. Postpublication strategies such as promotion of articles in social media and other content distribution networks promise to enhance discoverability, which may increase the chances that the most relevant paper is found and ultimately cited by scholars in the creation of their own work. However, how to target promotional campaigns to the right audience remains an open question, and there is currently limited evidence showing which strategies yield a positive effect on scholarly article impact, such as that reflected by citation counts.

One of the most widely studied strategies examined to increase discoverability, accessibility, and consequential citation counts of scholarly work is publishing content in open access (OA) journals. OA increases visibility and hence can have a positive effect on usage metrics such as the number of downloads [7], but the effect on citations is less clear and likely discipline dependent [8-12]. Since the first carefully designed multivariate observational study in 2006 showed a clear OA citation advantage even when adjusted for possible confounders [8,9,13], several large observational studies have found OA articles accrue between 50% to 200% more citations compared to closed access articles; this is referred to as the "open access citation advantage" (OACA) [8,14]. The OACA, however, has not been universally accepted [7,15,16]. For example, in 2011, Davis [16] completed a randomized controlled trial study showing that making articles free and open did not yield an increase in citation counts over a 3-year period relative to closed access articles. Part of the reason why the OACA has been found in observational data and not experimental data is likely due to residual confounders. Even though the original study suggesting an OA citation advantage was carefully adjusted for multiple confounders [8], selection bias cannot be fully adjusted for and is one of the biggest confounders identified in observational research examining the OACA [17]; more prominent authors are more likely to pay to publish OA articles, and if authors are more likely to provide access to their "highest quality" articles, then OA articles will have more citations than closed access articles [11,18]. Furthermore, as OA content becomes increasingly ubiquitous in the scholarly ecosystem, the OACA, if any, is likely to extinguish, as all articles share the same OA characteristic.

Aside from publishing research in an OA journal, other studies have suggested that the promotion of articles in social media platforms may be used to augment the page views and citations of articles, but this too remains highly controversial. For example, two recent studies [19,20] found that intensive social media promotion significantly increased page views of scholarly articles. A more recent study [21] found that an intensive social media promotion strategy yielded a citation advantage to promoted articles. However, the findings of the Luc et al [21] study have come under considerable scrutiny due to significant methodological flaws that suggests that social media promotion did not yield a citation benefit at all [22-24]. The most significant methodological flaw was the fact that the papers listed in the Luc et al [21] paper may have not matched those described by their methods. Similarly, two rigorous randomized controlled trials [25,26] found that the promotion of articles in social media did not yield any increase in article page views. These results refer to organic tweeting, and an unexamined question remains if promoted tweets, which measurably lead to more visits, also lead to more citations. Taken together, though there are mixed data as to whether social media promotion increases article page views, there are currently no replicated data to suggest that the strategy yields a citation advantage to promoted articles.

Our group previously published studies examining the extent to which the promotion of articles in a novel cross-publisher distribution platform (TrendMD) increases article page views, usage, and citation counts. In a 2014 study, TrendMD article promotion led to an 87% increase in page views relative to the control in a 4-week randomized controlled trial [27]. These data were replicated in in a 3-week crossover trial that found that promotion of articles in the TrendMD Network yielded a 30% and 49% weekly increase in page views relative to the control [28]. In 2017, we completed a 4-week randomized controlled trial that found articles randomized to TrendMD had a 77% increase in article saves on Mendeley relative to the control [29]. These findings were particularly significant because Mendeley saves are not only a robust measure of article usage, but the metric is also strongly correlated to future citations [30-34]. Building on these data, in 2019, we conducted a randomized controlled trial including 3200 articles published across 8 subject areas (eg, medicine, physics, business, and humanities) and found that promotion of articles over 6 months conferred an overall statistically significant 50% citation advantage to promoted articles relative to the control at 12 months [35]. TrendMD promotion increased citations for 3 of 8 disciplines tested, yielding the largest citation advantage to articles published in the subject areas of health, medical, and life sciences. Taken together, results of our studies suggest that TrendMD confers a short-term page view, usage, and citation advantage; however, we do not know whether the measured advantages persist over time or becomes diluted by other factors [36].

In this study, we conducted a follow-up analysis to our 12-month randomized controlled trial to determine whether the citation advantage persists at 36 months. We also sought to determine whether TrendMD's effect on citation counts were specific to particular disciplines or consistent across all disciplines. We hypothesized that the promotion of articles in TrendMD would yield a persistent citation advantage at 36 months. We further hypothesized that this effect would be seen across more disciplines than initially measured at 12 months.

Methods

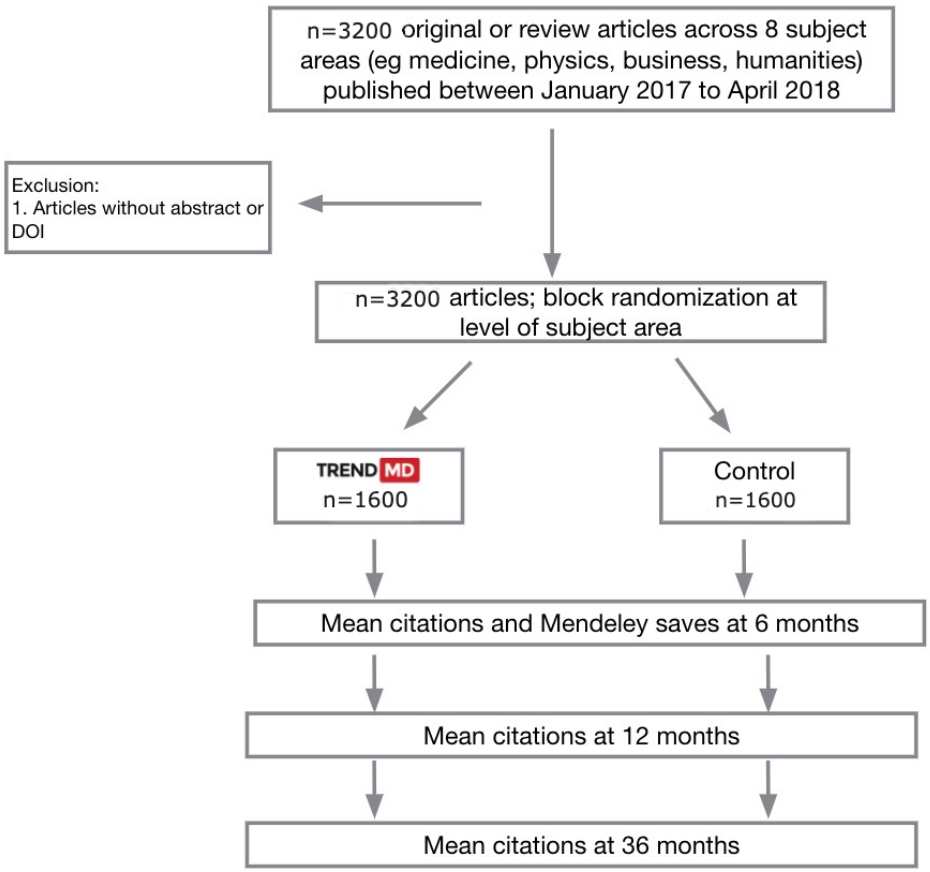
Summary

The majority of the Methods section included herein were copied verbatim from the Methods section we describe in our previously published 12-month randomized controlled trial [35].

We conducted a 36-month randomized controlled trial that included 3200 articles published in 64 peer-reviewed journals across 8 subject areas. The length of the study was 6 months for the intervention and an additional 30 months of observation for a total of 36 months. We published our initial findings at 12 months [35]. We measured citations at 6, 12, and 36 months. The subject areas/categories were selected based on the 8 categories listed in Google Scholar [37]. The categories selected were business economics and management; chemical and materials sciences; engineering and computer science; health and medical sciences; humanities, literature, and arts; life sciences and earth sciences; physics and mathematics; and social sciences. For each subject area, the top 20 journals ranked by Google Scholar's h-5 index were selected (Google Scholar only displays the top 20 journals in each subject area). Please see the supplementary material published in our 2019 study [35] for our rationale for using Google Scholar and the h-5 index in our selection criteria. Eight journals were then randomly selected with a random number generator from the top 20 in each subject area to be included in the study; we did this as opposed to selecting the top journals in each subject area so that our sample would include a randomized mixture of journals of high and lower impact in each subject area. Including both high and lower impact journals was important to our study because we wanted to mitigate the potential confounder that TrendMD promotion is only effective in high impact journals. Journals that were not indexed in Scopus or Web of Science were excluded from the study. Preprint servers such as ArXiv were also excluded from the study.

Starting from April 2018, 50 of the most recently published original articles or review articles in each journal were selected for inclusion in the study. Articles selected for inclusion were published online between January 2017 and April 2018; this includes early view articles. Articles were excluded if they did not contain an abstract or DOI. Block randomization using a random number generator at the subject level was used to randomize articles to either the control or the intervention arm of the study. For each subject area, 200 articles were randomized to the control, and 200 articles were randomized to the intervention. In total, 1600 articles were randomized to the control, and 1600 articles were randomized to the intervention. The overall study design is presented in Figure 1.

Figure 1. Overall study design.



Intervention

TrendMD [38] is a cross-publisher article recommendation and distribution platform that, as of May 2019, was embedded on over 4700 journals and websites from 300 publishers and seen by approximately 125 million readers per month. Roughly two-thirds of the TrendMD Network is related to scientific, technical, and medical (STM) publications; the other one-third is a relatively even split between social sciences, humanities,

and business publications. Participating publishers use TrendMD to distribute their published article links within the article recommendations displayed on articles within their journals (nonsponsored recommendations) or third-party journals within the TrendMD Network (sponsored recommendations; please see Figures 2 and 3). TrendMD's content distribution model is benchmarked to similar services in the consumer web, where the leading networks Outbrain [39] and Taboola [40] generate the "From the web" and "You may like" recommendations seen alongside the content on many popular websites like CNN or BBC [29,35] (please see Figure 4 for reference).

The intervention consisted of the promotion of 1600 articles in the TrendMD Network for 6 months, between May 1, 2018, and November 2, 2018. Articles included in the TrendMD Network are displayed as recommended article links. Links to articles randomized to TrendMD were displayed as sponsored recommended links on publications participating in the TrendMD Network. There was an average of 4300 participating journals and 121 million readers per month during the course of the study. The frequency of sponsored article link placements is determined by a relevancy score based on the following: relatedness (ie, keyword overlap), collaborative filtering (similar to Amazon's "people who bought this item also bought that item"), and user clickstream analysis (the Netflix approach, basing recommendations on the users' interests expressed through their online history) [27-29,35]. As a result of the relevancy scoring system, some articles randomized to TrendMD were both seen more often (ie, accrued more link impressions) and clicked on more frequently than others in the TrendMD Network.

Figure 2. TrendMD: sponsored versus nonsponsored links.

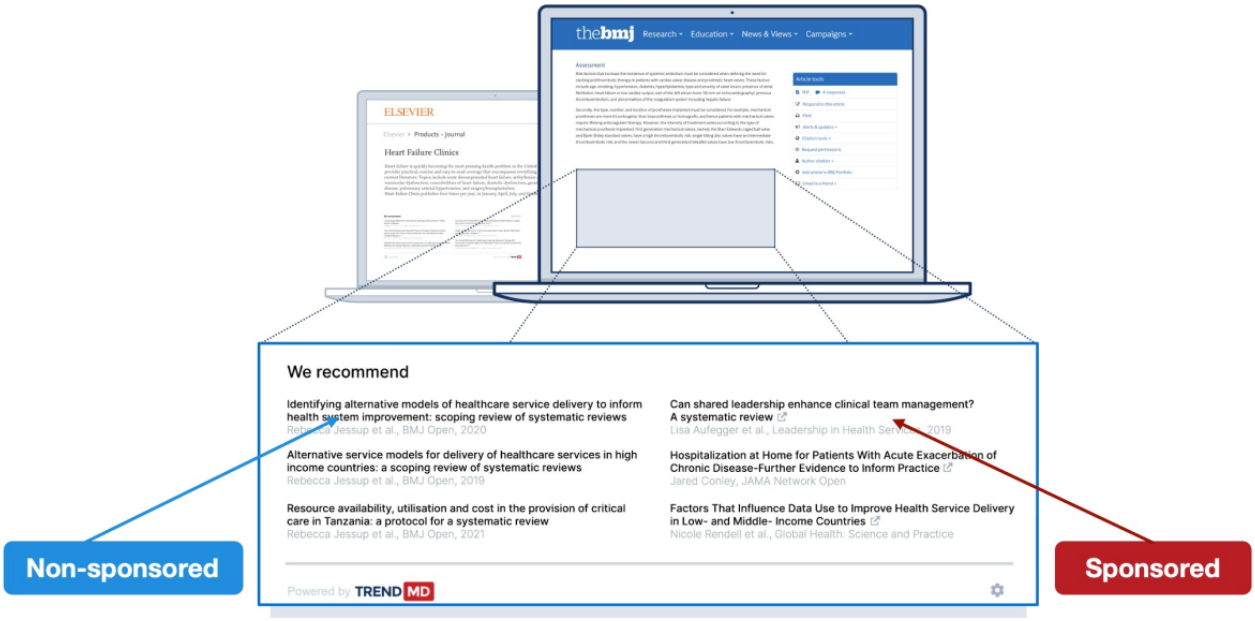
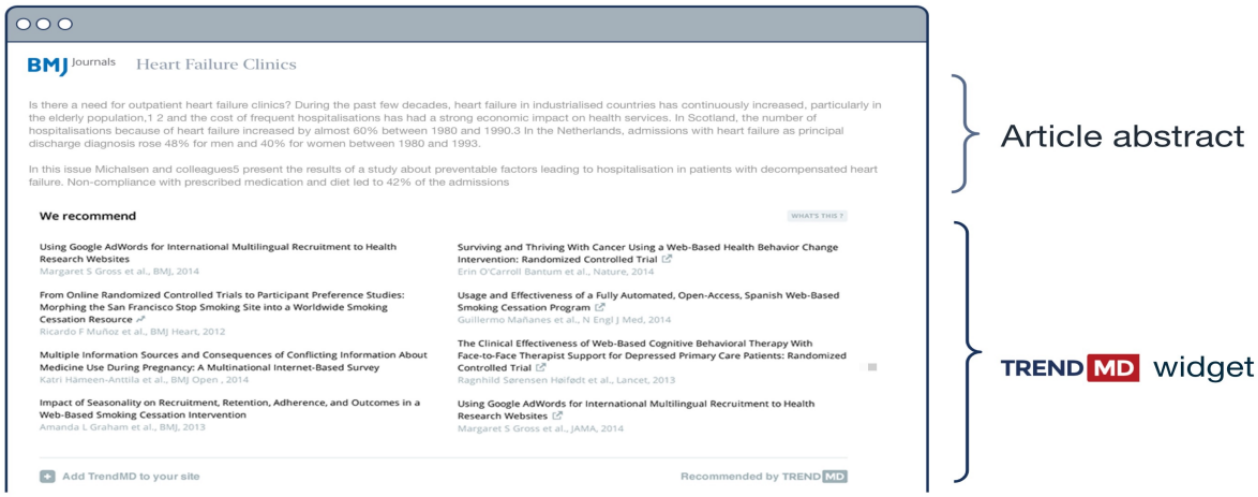


Figure 3. How the TrendMD Network works.

Figure 4. TrendMD widget display.



The sponsored links are displayed in the TrendMD Network as long as they are relevant and the advertiser account balance is greater than US $0. An account in TrendMD was created for this study. The 1600 articles randomized to TrendMD received a maximum total budget of US $9600 at a cost-per-click of US $0.10 for 96,000 sponsored TrendMD clicks. The account was allowed to spend up to a maximum for US $1600 per month, or 16,000 clicks per month, over the 6-month study period for a total of 96,000 clicks. The actual amount spent by the account was US $1600 per month; all clicks were delivered each month throughout the 6-month study. A summary of how TrendMD works and the outcomes measured can be seen in Figure 3.

Control

Articles randomized to the control (n=1600) received no promotion in the TrendMD Network. Articles randomized to the control received traffic by organic means (eg, Google, Google Scholar, or PubMed) and other means implemented by publishers or authors of content outside the context of this study.

Primary Outcome

The primary outcome of our study is the mean citation counts for articles randomized to TrendMD compared to the control at 36 months. Article citation counts at 36 months were extracted through the Scopus application programming interface (API) on March 30, 2021.

Secondary Outcomes

Thirty-six–Month Analysis

Mean citation counts after 36 months were compared for articles randomized to TrendMD versus control for each of the 8 subject areas. This analysis was completed to assess whether the effects of TrendMD promotion were discipline specific. Article citation counts were extracted from the Scopus API on March 30, 2021.

Twelve- and Six-Month Analysis

Mean citation counts after 12 and 6 months were compared for articles randomized to TrendMD versus the control. Article citation counts were extracted from the Scopus API on May 2, 2019, for the 12-month data and on November 2, 2018, for the 6-month data. These data were also published in our 2019 paper and included here for convenience to readers [35].

Statistical Analysis

We performed an a priori power calculation to determine the necessary sample size (n=1600 in each arm of the study) to detect differences in our primary outcome of mean citation counts between groups at 12 months. Please see our previously published paper for how we determined the required sample size in each arm of the study [35]. The study was not powered to detect differences in citation counts at 6 or 36 months nor was it powered to detect differences across the subject area–level comparisons at any of the time intervals. To power a study to detect differences at the subject level, each subject would have required between 1000 to 3000 articles in each arm of the study, which was not feasible for us to conduct for budgetary reasons.

Baseline characteristics of articles at the start of the study were tabulated and compared across randomized arms of the study. We categorized articles by subject area, access type (closed vs OA), Journal Impact Factor, and citations and Mendeley saves. Both the primary and secondary outcomes were analyzed with the two-sample *t* test on log-transformed data (1 + x). In the event that mean differences were statistically significant, we calculated effect sizes using Cohen *d* [41] on log-transformed data. Cohen *d* is defined as the difference between two means divided by a SD for the data. Lastly, we performed a stepwise and backward multivariate ordinary least squares (linear) regression analysis to determine the predictors of citations at 6, 12, and 36 months; we used log-transformed dependent variables in the model. A significance value of 0.1 was used for the removal of variables in our stepwise regression model. Journal Impact Factor, access type (ie, OA vs closed), baseline Mendeley saves and citations, and TrendMD clicks (ie, clicks on promoted article links) and impressions (ie, display of promoted article links) were covariates in the regression model; they were selected as covariates because each of them has known independent effects on citations (eg, Journal Impact Factor is a predictor of citation counts) and do not have issues of multicollinearity. We did not collect data on all possible predictors of citations (eg, number of authors or international collaboration) in our regression model, as this was out of scope

to our analysis; the primary goal of our regression model was to determine if TrendMD clicks and impressions were independent predictors of our primary outcome of citations at 36 months. All regressions adjusted the SEs for clustering of citations using Huber-White SEs; this corrected for heteroscedasticity [42]. Residuals were analyzed for normality using the Kolmogorov-Smirnov test; we also analyzed the residuals using skewness and kurtosis statistical values and compared to SE values.

A 2-tailed $P<.05$ was considered statistically significant. To mitigate the type I error rate, the Bonferroni correction method was used to control for multiple comparisons made for the 8 disciplinary differences in mean citation counts [43]. A 2-tailed $P<.006$ was considered to be statistically significant for mean differences in Mendeley reader and citation counts across subject areas. Arithmetic mean values for Mendeley saves and citation counts are shown with 95% CIs on non–log-transformed data unless otherwise specified. Tests for normality were included in the model. SPSS version 25 (IBM Corp) was used to complete the statistical analyses.

Results

Baseline Characteristics

The following was published verbatim in our previously published randomized controlled trial [35]; these data and discussion are included here for convenience of the reader. Overall, 3200 articles were randomized: 1600 to the TrendMD arm and 1600 to the control arm. The Kolmogorov-Smirnov test of the 6-month Mendeley saves ($P=.61$) and citation count ($P=.13$) data retained the null hypothesis that the distributions were log-normal within the control and TrendMD arms. Tables 1 and 2 show the baseline characteristics of articles randomized to TrendMD versus the control; no statistical tests were used to compare any metrics between groups at baseline according to the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines [44].

Table 1. Overall baseline characteristics [35].

Intervention	Articles, n	Open access, n	Journal Impact Fact, mean (SD)	Mendeley saves, mean (SD)	Citation count, mean (SD)
Control	1600	88	14.77 (15.87)	23.37 (36.33)	0.86 (2.98)
TrendMD	1600	92	15.41 (16.56)	23.79 (39.34)	1.04 (3.37)

Table 2. Baseline characteristics by subject area [35].

Category and intervention	Open access, n	Journal Impact Factor, mean (SD)	Mendeley saves, mean (SD)	Citation count, mean (SD)
Business, economics, and management				
Control	0	5.28 (1.59)	33.99 (43.69)	1.02 (1.60)
TrendMD	0	5.53 (1.55)	35.38 (35.60)	1.24 (2.26)
Chemical and materials sciences				
Control	1	26.53 (12.20)	23.92 (26.95)	1.53 (7.27)
TrendMD	1	31.37 (15.89)	25.97 (43.57)	1.58 (6.31)
Engineering and computer science				
Control	3	14.48 (9.74)	10.44 (16.27)	0.29 (0.65)
TrendMD	2	14.82 (9.93)	10.27 (16.83)	0.35 (0.97)
Health and medical sciences				
Control	11	35.98 (22.48)	37.98 (52.55)	1.93 (2.29)
TrendMD	7	34.42 (21.87)	38.04 (62.8)	3.1 (5.55)
Humanities, literature, and arts				
Control	2	2.15 (0.86)	13.39 (17.36)	1.06 (2.08)
TrendMD	2	2.15 (0.87)	13.13 (15.97)	0.95 (2.09)
Life sciences and earth sciences				
Control	52	21.57 (13.8)	43.33 (53.73)	0.37 (0.74)
TrendMD	51	22.76 (14.42)	44.13 (52.87)	0.48 (0.85)
Physics and mathematics				
Control	10	7.67 (5.86)	8.90 (14.33)	0.33 (1.09)
TrendMD	11	7.74 (5.85)	9.04 (18.12)	0.31 (1.29)
Social sciences				
Control	9	4.52 (1.08)	15.01 (16.56)	0.36 (1.16)
TrendMD	18	4.51 (1.06)	14.39 (17.27)	0.35 (1.17)

Primary Outcome

Articles randomized to TrendMD (n=1600) showed a 28% increase in mean citations relative to the control at 36 months (Figure 5). The mean citations for articles randomized to TrendMD was 48.05 (SD 113.51), compared to 37.53 (SD 77.36) for articles randomized to the control. The difference in mean citations for TrendMD articles versus the control was 10.52 (95% CI 3.79-17.25) and was statistically significant (P=.001). The effect size of TrendMD on citations at 36 months was small (Cohen d 0.11; Table 3). The cumulative distribution of article citations at 36 months is shown in Figure 6. The fact that the TrendMD cumulative distribution curve is shifted to the right relative to the control indicates the promotion of increased citation rates across the entire distribution of articles; the effect was not limited to just a few outlying articles.

Figure 5. Mean citation count over 36 months: TrendMD versus control.

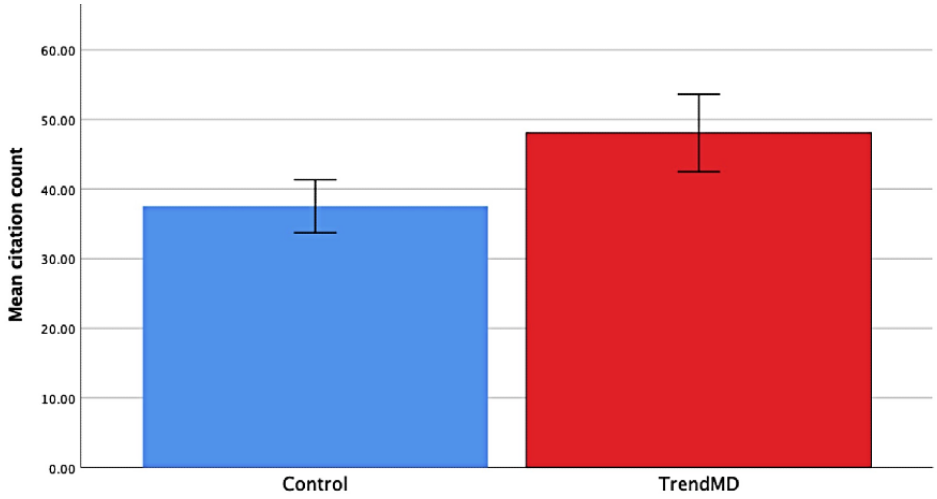
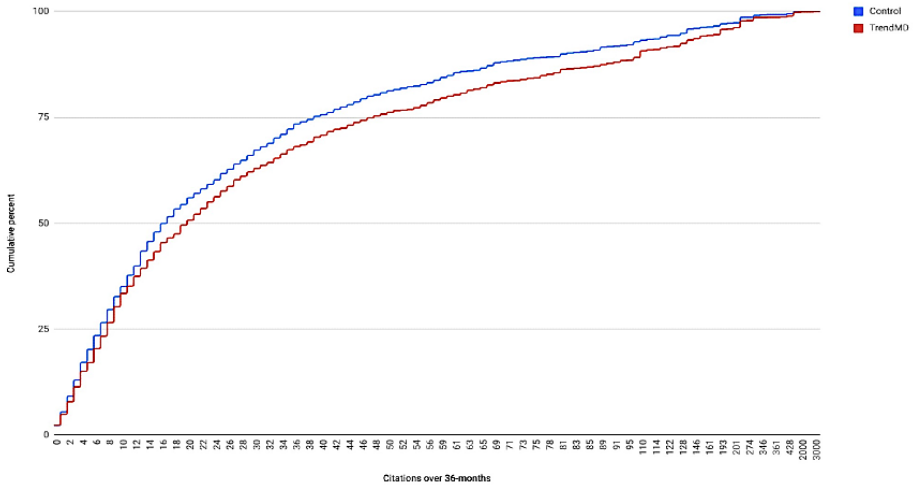


Table 3. Citation counts at 36, 12, and 6 months.

	36 months		12 months		6 months	
	Control	TrendMD	Control	TrendMD	Control	TrendMD
Citations, mean (SD)	37.53 (77.36)	48.05 (113.51)	10.10 (19.09)	15.16 (40.37)	5.12 (10.65)	6.18 (16.10)
Mean difference in citations (95% CI)	N/A[a]	10.52 (3.79-17.25)	N/A	5.06 (2.87-7.25)	N/A	1.06 (0.12-2.01)
P value	N/A	.001	N/A	<.001	N/A	.005
Cohen *d*	N/A	0.11	N/A	0.16	N/A	0.10

[a]N/A: not applicable.

Figure 6. Cumulative distribution of citations over 36 months: TrendMD versus control.



Secondary Outcomes

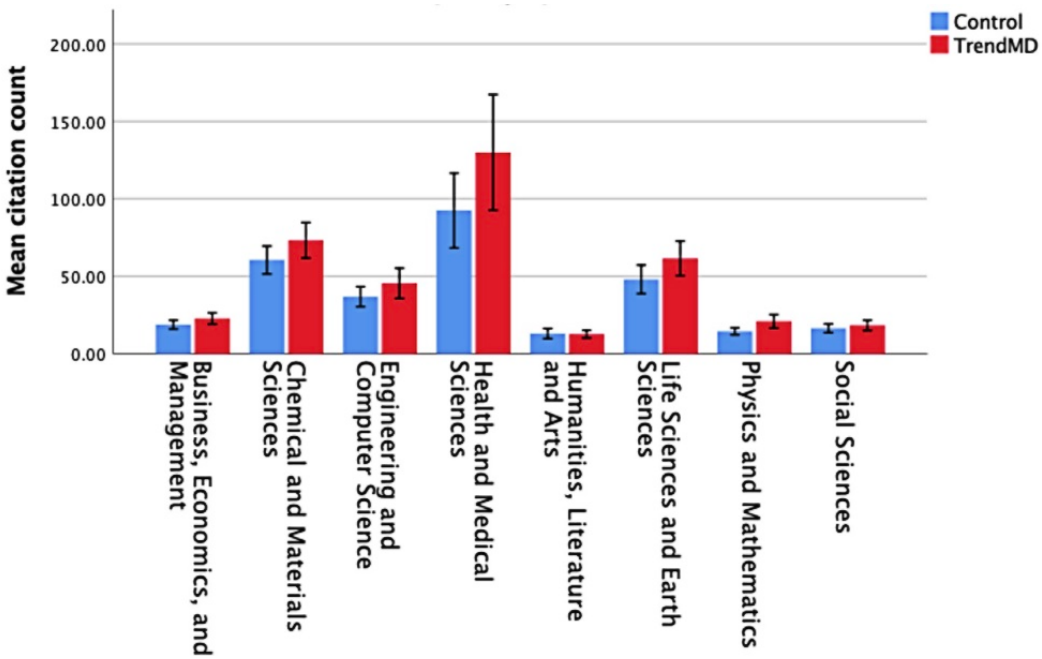
Subject Area Differences in Mean Citations at 36 Months

At 36 months, TrendMD was not found to yield statistically significant increases in citation counts relative to the control in any of the individual subject areas. The largest relative difference in citation counts compared to the control was found to be in the subject area of health and medical sciences (41%); however, it was not found to be statistically significant (see Table 4 and Figure 7 for a breakdown of mean differences for each subject area).

Table 4. Citations at 36 months by subject area.

Category and intervention	Articles, n	Citations, mean (SD)	Mean difference (95% CI)	*P* value
Business, economics, and management			3.95 (–0.66 to 8.56)	.24
Control	200	18.75 (20.45)		
TrendMD	200	22.70 (26.14)		
Chemical and materials sciences			12.68 (–1.84 to 27.20)	.18
Control	200	60.46 (64.48)		
TrendMD	200	73.14 (82.17)		
Engineering and computer science			8.61 (–3.07 to 20.29)	.40
Control	200	36.80 (46.04)		
TrendMD	200	45.41 (70.31)		
Health and medical sciences			37.49 (–6.91 to 81.89)	.08
Control	200	92.43 (173.73)		
TrendMD	200	129.92 (267.99)		
Humanities, literature, and arts			–0.35 (–3.73 to 4.43)	.89
Control	200	12.98 (23.63)		
TrendMD	200	12.63 (17.40)		
Life sciences and earth sciences			13.51 (–0.91 to 27.93)	.03
Control	200	47.97 (66.14)		
TrendMD	200	61.48 (79.95)		
Physics and mathematics			6.43 (–1.53 to 11.33)	.12
Control	200	14.46 (16.00)		
TrendMD	200	20.89 (31.38)		
Social sciences			1.90 (–2.4 to 6.20)	.05
Control	200	16.39 (20.13)		
TrendMD	200	18.29 (23.52)		

Figure 7. Mean citation counts by subject area over 36 months: TrendMD versus control.



Multivariate Regression Analysis

Overall, our multivariate regression models were a good fit to predict citations in all articles at 6, 12, and 36 months. The Kolmogorov-Smirnov test of the residuals (*P*=.21) were normal; values for skewness and kurtosis in the residuals were both less than 1 SE, which suggests that the residual values were not significantly different from the expected value of zero for a normal distribution. Our models predicted 41%, 43%, and 44% of the variation in citation counts at 6, 12, and 36 months, respectively. Clicks driven by TrendMD were found to be an independent predictor of citations; articles that received the greatest number of clicks from TrendMD had the largest citation advantage at 36 months (see Table 5 for the standardized beta coefficients for each covariate in the model).

Table 5. Multivariate regression model: standardized beta coefficients at 36, 12, and 6 months.[a]

Standardized beta coefficients	36 months	12 months	6 months
Baseline citations	0.286	0.329	0.323
Baseline Mendeley saves	0.209	0.216	0.163
Journal Impact Factor	0.207	0.175	0.222
Access type	NS[b] (0.664)[c]	NS (0.847)	NS (0.721)
TrendMD clicks	0.266	0.182	0.153
TrendMD impressions	NS (0.342)	0.063	0.073

[a]All other variables included in the regression model were significant at *P*<.001.

[b]NS: nonsignificant.

[c]Significance values for nonsignificant variables excluded in the model are included in brackets.

Discussion

To the best of our knowledge, this was the first randomized controlled trial to demonstrate how relatively brief postpublication promotion of peer-reviewed articles over a 6-month period can be used to persistently increase citation rates of those articles after 36 months. The overall citation advantage conferred by article promotion in the TrendMD Network was 28% relative to the control at 36 months, which dissipated from a 50% citation advantage at 12 months. Despite the overall citation advantage observed at 36 months, we did not find a statistically significant increase to citation counts within individual subject areas included in this study. There were, however, larger, albeit nonstatistically significant differences in citation counts measured in certain subjects versus others. For example, health and medical sciences articles saw larger citation increases than articles published in humanities, literature, and arts journals. One possible explanation for this could be because two-thirds of the TrendMD Network are STM journals; therefore, the effect size of promotion is likely to be larger for STM articles rather than humanities-related articles. The other possible explanation for the smaller citation advantages in some subject areas is because of differing publication cycle lengths, which have a direct effect on the speed in which citations accrue overtime. Articles in the fields of medicine, life sciences, and physics, for example, are typically published faster in comparison to articles in the humanities and social sciences [2].

Notwithstanding, our study was not powered to detect differences across the subject area–level comparisons; therefore, the negative findings across individual subject areas are likely due to the small sample sizes and resulting type II errors [45]. Given that we found an overall citation advantage at 36 months and found a Mendeley save advantage in 7 of the 8 subject areas at 6 months, it is reasonable to conclude that studies using larger sample sizes within subject areas over periods longer than 36 months may have shown an increase in citations. Future studies using larger sample sizes within subject areas over longer periods of time are needed to determine whether, and to what degree, promotion of articles in the TrendMD Network confers a citation advantage to individual subject areas.

We initially hypothesized that the overall 50% citation advantage observed at 12 months would increase over time due to the time it takes citations to accrue from manuscripts passing through peer review and onto publication [2]. However, our data suggests the citation advantage conferred to promoted articles dissipated between 12 and 36 months, which could be due to multiple factors affecting article citation rates, such as author credentials, subject matter, and cited references [36]. We do not know the precise timing of the peak of the citation advantage because we only measured citation counts at 6, 12, and 36 months. Future studies are warranted that measure citation rates at more frequent intervals to discern when the citation advantage conferred to promoted articles are at its maximum.

Based on these data, we can speculate on the mechanism in which TrendMD conferred a sustained citation advantage to promoted articles at 36 months. Readers clicking on promoted article links displayed in the widget that they may have not otherwise discovered, saved these articles to their Mendeley reference libraries, and cited these articles while creating their own scholarly work. This pathway is evidenced by the fact that page views (ie, clicks) driven by TrendMD were an independent predictor of both Mendeley saves at 6 months and citations at 6, 12, and 36 months (Table 5).

The findings of our study significantly add to the limited corpus of literature examining the efficacy of strategies to augment citation counts of peer-reviewed research. Though publishing research in OA journals is the most widely studied strategy to

augment citation counts, the extent to which the OACA exists or is confounded by selection bias remains unknown [7,11,12,15,16]. In addition, even if the OACA did exist at some point in time in the past, the more that OA content becomes the standard, the less likely the strategy will yield citation benefits as the effect will be ubiquitous across all articles. Other strategies such as the promotion of peer-reviewed literature in social media platforms for the purpose of enhancing citation counts have similarly yielded inconclusive and, at times, conflicting results [21-23,25,26]. These data presented in our study address a pressing unmet need of authors, publishers, and funders for evidence-based strategies that can be used to enhance discoverability by the right targeted audience, which in turn augments the usage and citations counts of peer-reviewed content.

The research presented here has several strengths. First, the outcome of citation counts is unbiased and objective, increasing the reproducibility of the results. Second, we used a rigorous randomized controlled trial design, which minimizes the likelihood of bias and confounding. Third, our sample size was large, and our study was adequately powered for outcomes of differences in mean citations.

There are, however, several limitations to this research. First, all authors have a conflict of interest with the results presented as creators (PK and GE) or employees (TB) of TrendMD. Risk of bias, however, was mitigated by the randomized controlled trial design. Second, although our study was adequately powered to detect mean differences in citation counts across all articles at 12 months, our study was not adequately powered to detect mean differences in citation counts between disciplines. Third, our inclusion criteria of randomly selecting articles published in 8 out of the top 20 journals with the highest h-5 index in Google Scholar categories may make our results less generalizable to articles published in journals with lower Impact Factors. We attempted to mitigate this possible limitation by randomly selecting 8 of the top 20 journals in each subject. Future studies are needed to determine whether TrendMD still confers an enduring citation advantage to articles published in lower Impact Factor journals. Another limitation and question that stems from these data is whether longer periods of promotion beyond 6 months would yield larger and more enduring benefits to citation counts. Future studies using promotion periods longer than 6 months are needed to determine whether there is a dose-dependent relationship between the length of the article promotion period and the magnitude of citation advantage conferred. Our group has previously completed a study showing that there is indeed a dose-dependent relationship between the intensity of article promotion and page views [28]; however, the degree to which the increase in page views leads to a citation advantage is not known. It is also possible that a longer promotion period of articles may saturate over time as readers are presented with the same article links; future studies are needed to determine whether prolonged promotion of articles lead to diminishing rates of return for citations. Lastly, the number and type of publishers participating in the TrendMD Network may change over time, which may affect the reproducibility of our findings [35]. Past studies indicate that the efficacy of TrendMD promotion is dependent on the number and type of publishers participating in the Network [28,29]. In general, replicated data indicates that the more publishers across subject areas using TrendMD the greater the efficacy of the channel to confer benefits to article visibility and usage [35]. If publishers stopped using TrendMD, the channel is unlikely to be as effective at augmenting citation counts as described in this study. More generally, there are, of course, limitations to citation analyses; citations only reflect activity in academia, and the usefulness of citations as indicators varies greatly between fields [46].

Despite the limitations, this study demonstrates that the promotion of articles in a cross-publisher online distribution channel (TrendMD) over a 6-month period can be used to persistently increase citations of articles after 36 months. Though we did not find a statistically significant increase in citation counts within individual subject areas at 36 months, this was likely due to small sample sizes and insufficient power, which resulted in type II errors. The overall citation advantage conferred to promoted articles was observed at 6 months, appeared to peak at 12 months, and endured, albeit to a lower level relative to the control at 36 months.

Acknowledgments

This research was supported in part by two investment grants made by the Ontario Centres of Excellence (OCE), a Canadian government-funding program. The grants were OCE Market Readiness CC (Grant 22292) and OCE Market Readiness CB (Grant 23811). We also want to thank members of the TrendMD team for help with data acquisition.

Conflicts of Interest

PK is a cofounder and former Chief Executive Officer (CEO) of TrendMD Inc. TB is a full-time employee of TrendMD Inc. GE is a cofounder and former Chief Science Officer of TrendMD Inc. GE and PK currently hold no equity in TrendMD. GE is also founder and CEO of JMIR Publications, where TrendMD was initially tested and developed. The reviewers and editors of this paper were not aware of GE's coauthorship during the peer-review process. JMIR Publications continues to offer TrendMD services to its authors and receives a commission.

This randomized study was not registered, explained by authors with the reason that the study does not require registration due to its nature. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the study is a follow-up on an initial study involving peer-reviewed articles. It does not involve the randomization of people into groups and does not target any behavioural outcomes.

Multimedia Appendix 1
CONSORT non-eHealth checklist.
[PDF File (Adobe PDF File), 65 KB - jmir_v23i12e34051_app1.pdf]

References

1. Aksnes DW, Langfeldt L, Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. SAGE Open 2019 Feb 07;9(1):215824401982957. [doi: 10.1177/2158244019829575]
2. Johnson R, Watkinson A, Mabe M. The STM Report. White Paper Sep 24. Report No 1. STM. 2018 Jan 01. URL: https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf [accessed 2021-12-02]
3. Tenopir C, Christian L, Kaufman J. Seeking, reading, and use of scholarly articles: an international study of perceptions and behavior of researchers. Publications 2019 Mar 06;7(1):18-25. [doi: 10.3390/publications7010018]
4. Van Noorden R. Scientists may be reaching a peak in reading habits. Nature 2014 Feb 3:15-17. [doi: 10.1038/nature.2014.14658]
5. Van Noorden R. The science that's never been cited. Nature 2017 Dec 14;552(7684):162-164. [doi: 10.1038/d41586-017-08404-0] [Medline: 29239363]
6. Chu JSG, Evans JA. Slowed canonical progress in large fields of science. Proc Natl Acad Sci U S A 2021 Oct 12;118(41):e2021636118 [FREE Full text] [doi: 10.1073/pnas.2021636118] [Medline: 34607941]
7. Davis PM, Lewenstein BV, Simon DH, Booth JG, Connolly MJL. Open access publishing, article downloads, and citations: randomised controlled trial. BMJ 2008 Jul 31;337:a568 [FREE Full text] [doi: 10.1136/bmj.a568] [Medline: 18669565]
8. Eysenbach G. Citation advantage of open access articles. PLoS Biol 2006 May;4(5):e157 [FREE Full text] [doi: 10.1371/journal.pbio.0040157] [Medline: 16683865]
9. Eysenbach G. The open access advantage. J Med Internet Res 2006 May 15;8(2):e8 [FREE Full text] [doi: 10.2196/jmir.8.2.e8] [Medline: 16867971]
10. Björk BC, Solomon D. Open access versus subscription journals: a comparison of scientific impact. BMC Med 2012 Jul 17;10:73 [FREE Full text] [doi: 10.1186/1741-7015-10-73] [Medline: 22805105]
11. Copiello S. The open access citation premium may depend on the openness and inclusiveness of the indexing database, but the relationship is controversial because it is ambiguous where the open access boundary lies. Scientometrics 2019 Oct 10;121(2):995-1018. [doi: 10.1007/s11192-019-03221-w]
12. Basson I, Blanckenberg JP, Prozesky H. Do open access journal articles experience a citation advantage? Results and methodological reflections of an application of multiple measures to an analysis by WoS subject areas. Scientometrics 2020 Oct 17;126(1):459-484. [doi: 10.1007/s11192-020-03734-9]
13. MacCallum CJ, Parthasarathy H. Open access increases citation rate. PLoS Biol 2006 May;4(5):e176 [FREE Full text] [doi: 10.1371/journal.pbio.0040176] [Medline: 16683866]
14. Ottaviani J. The post-embargo open access citation advantage: it exists (probably), its modest (usually), and the rich get richer (of course). PLoS One 2016;11(8):e0159614 [FREE Full text] [doi: 10.1371/journal.pone.0159614] [Medline: 27548723]
15. Anderson R. Is rational discussion of open access possible? Insights 2014 Jul 7;27(2):171-180. [doi: 10.1629/2048-7754.162]
16. Davis PM. Open access, readership, citations: a randomized controlled trial of scientific journal publishing. FASEB J 2011 Jul;25(7):2129-2134. [doi: 10.1096/fj.11-183988] [Medline: 21450907]
17. Craig I, Plume A, McVeigh M, Pringle J, Amin M. Do open access articles have greater citation impact? A critical review of the literature. J Informetrics 2007 Jul;1(3):239-248. [doi: 10.1016/j.joi.2007.04.001]
18. Kurtz MJ, Eichhorn G, Accomazzi A, Grant C, Demleitner M, Henneken E, et al. The effect of use and access on citations. Inf Processing Manage 2005 Dec;41(6):1395-1402. [doi: 10.1016/j.ipm.2005.03.010]
19. Hawkins CM, Hunter M, Kolenic GE, Carlos RC. Social media and peer-reviewed medical journal readership: a randomized prospective controlled trial. J Am Coll Radiol 2017 May;14(5):596-602. [doi: 10.1016/j.jacr.2016.12.024] [Medline: 28268163]
20. Trueger NS, Bokarius AV, Carroll S, April MD, Thoma B. Impact of a physician-led social media sharing program on a medical journal's web traffic. J Am Coll Radiol 2018 Jan;15(1 Pt B):184-189. [doi: 10.1016/j.jacr.2017.09.035] [Medline: 29122507]
21. Luc JGY, 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 2021 Jan;111(1):296-300. [doi: 10.1016/j.athoracsur.2020.04.065] [Medline: 32504611]
22. Davis P. Reanalysis of tweeting study yields no citation benefit. The Scholarly Kitchen. URL: https://scholarlykitchen.sspnet.org/2020/07/13/tweeting-study-yields-no-benefit/ [accessed 2021-12-02]
23. Davis P. Tweeting-citations authors speak, finally. The Scholarly Kitchen. URL: https://scholarlykitchen.sspnet.org/2020/08/03/tweeting-citations-authors-speak-finally/ [accessed 2021-12-02]
24. Luc JGY, Varghese TK, Antonoff MB. We stand by our data: a call for professional scholarly discourse. Ann Thorac Surg 2021 Mar;111(3):1095-1097. [doi: 10.1016/j.athoracsur.2020.07.007] [Medline: 32707195]

25. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from Circulation. Circulation 2015 Jan 06;131(1):28-33. [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
26. 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):e003088 [FREE Full text] [doi: 10.1161/JAHA.115.003088] [Medline: 27121850]
27. Kudlow P, Rutledge A, Eysenbach G. TrendMD: reaching larger, more targeted audiences by distributing scholarly content online. Editors Bull 2015 Jun 08;10(1):11-15. [doi: 10.1080/17521742.2015.1031965]
28. Kudlow P, Rutledge A, Shachak A, McIntyre RS, Eysenbach G. TrendMD: helping scholarly content providers reach larger and more targeted audiences. Learned Publishing 2016 Feb 16;29(2):125-129. [doi: 10.1002/leap.1014]
29. Kudlow P, Cockerill M, Toccalino D, Dziadyk DB, Rutledge A, Shachak A, et al. Online distribution channel increases article usage on Mendeley: a randomized controlled trial. Scientometrics 2017;112(3):1537-1556 [FREE Full text] [doi: 10.1007/s11192-017-2438-3] [Medline: 28804178]
30. Aduku KJ, Thelwall M, Kousha K. Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of computer science and engineering. Scientometrics 2017 Apr 13;112(1):573-581. [doi: 10.1007/s11192-017-2367-1]
31. Li X, Thelwall M, Giustini D. Validating online reference managers for scholarly impact measurement. Scientometrics 2011 Dec 21;91(2):461-471. [doi: 10.1007/s11192-011-0580-x]
32. Li X, Thelwall M. F1000, Mendeley and traditional bibliometric indicators. 2012 Presented at: Proceedings of the 17th International Conference of Bibliometrics; 2012; Valencia, Spain.
33. Thelwall M, Nevill T. Could scientists use Altmetric.com scores to predict longer term citation counts? J Informetrics 2018 Feb 15;12(1):237-248. [doi: 10.1016/j.joi.2018.01.008]
34. Thelwall M, Sud P. Mendeley readership counts: an investigation of temporal and disciplinary differences. J Assoc Inf Sci Technol 2015 Jun 05;67(12):3036-3050. [doi: 10.1002/asi.23559]
35. Kudlow P, Bissky Dziadyk D, Rutledge A, Shachak A, Eysenbach G. The citation advantage of promoted articles in a cross‐publisher distribution platform: a 12‐month randomized controlled trial. J Assoc Inf Sci Technol 2019 Dec 23;71(10):1257-1274. [doi: 10.1002/asi.24330]
36. Onodera N, Yoshikane F. Factors affecting citation rates of research articles. J Assoc Inf Sci Technol 2014 May 21;66(4):739-764. [doi: 10.1002/asi.23209]
37. Top publications 2021. Google. URL: https://scholar.google.ca/citations?view_op=top_venues&hl=en [accessed 2021-12-02]
38. TrendMD. URL: https://trendmd.com/ [accessed 2021-12-02]
39. Outbrain. URL: https://www.outbrain.com/ [accessed 2021-12-02]
40. Taboola. URL: https://www.taboola.com/ [accessed 2021-12-02]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988.
42. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980 May;48(4):817. [doi: 10.2307/1912934]
43. Chen S, Feng Z, Yi X. A general introduction to adjustment for multiple comparisons. J Thorac Dis 2017 Jun;9(6):1725-1729. [doi: 10.21037/jtd.2017.05.34] [Medline: 28740688]
44. de Boer MR, Waterlander WE, Kuijper LD, Steenhuis IH, Twisk JW. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015 Jan 24;12:4 [FREE Full text] [doi: 10.1186/s12966-015-0162-z] [Medline: 25616598]
45. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med 1978 Sep 28;299(13):690-694. [doi: 10.1056/NEJM197809282991304] [Medline: 355881]
46. MacRoberts MH, MacRoberts BR. Problems of citation analysis. Scientometrics 1996 Jul;36(3):435-444. [doi: 10.1007/bf02129604]

Abbreviations

API: application programming interface
CEO: Chief Executive Officer
CONSORT: Consolidated Standards of Reporting Trials
OA: open access
OACA: open access citation advantage
OCE: Ontario Centres of Excellence
STM: scientific, technical, and medical

Edited by R Kukafka; submitted 05.10.21; peer-reviewed by M Thelwall; comments to author 27.10.21; accepted 21.11.21; published 10.12.21.

Please cite as:
Kudlow P, Brown T, Eysenbach G
Citation Advantage of Promoted Articles in a Cross-Publisher Distribution Platform: 36-Month Follow-up to a Randomized Controlled Trial
J Med Internet Res 2021;23(12):e34051
URL: https://www.jmir.org/2021/12/e34051
doi:10.2196/34051
PMID:34890350

Original Paper

Predicting New Daily COVID-19 Cases and Deaths Using Search Engine Query Data in South Korea From 2020 to 2021: Infodemiology Study

Atina Husnayain[1], MPH; Eunha Shim[2], PhD; Anis Fuad[3], DEA; Emily Chia-Yu Su[1,4], PhD

[1]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[2]Department of Mathematics, Soongsil University, Seoul, Republic of Korea

[3]Department of Biostatistics, Epidemiology, and Population Health, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

[4]Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan

Corresponding Author:
Emily Chia-Yu Su, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
172-1 Keelung Rd, Sec 2
Taipei, 106
Taiwan
Phone: 886 266382736 ext 1515
Email: emilysu@tmu.edu.tw

Abstract

Background: Given the ongoing COVID-19 pandemic situation, accurate predictions could greatly help in the health resource management for future waves. However, as a new entity, COVID-19's disease dynamics seemed difficult to predict. External factors, such as internet search data, need to be included in the models to increase their accuracy. However, it remains unclear whether incorporating online search volumes into models leads to better predictive performances for long-term prediction.

Objective: The aim of this study was to analyze whether search engine query data are important variables that should be included in the models predicting new daily COVID-19 cases and deaths in short- and long-term periods.

Methods: We used country-level case-related data, NAVER search volumes, and mobility data obtained from Google and Apple for the period of January 20, 2020, to July 31, 2021, in South Korea. Data were aggregated into four subsets: 3, 6, 12, and 18 months after the first case was reported. The first 80% of the data in all subsets were used as the training set, and the remaining data served as the test set. Generalized linear models (GLMs) with normal, Poisson, and negative binomial distribution were developed, along with linear regression (LR) models with lasso, adaptive lasso, and elastic net regularization. Root mean square error values were defined as a loss function and were used to assess the performance of the models. All analyses and visualizations were conducted in SAS Studio, which is part of the SAS OnDemand for Academics.

Results: GLMs with different types of distribution functions may have been beneficial in predicting new daily COVID-19 cases and deaths in the early stages of the outbreak. Over longer periods, as the distribution of cases and deaths became more normally distributed, LR models with regularization may have outperformed the GLMs. This study also found that models performed better when predicting new daily deaths compared to new daily cases. In addition, an evaluation of feature effects in the models showed that NAVER search volumes were useful variables in predicting new daily COVID-19 cases, particularly in the first 6 months of the outbreak. Searches related to logistical needs, particularly for "thermometer" and "mask strap," showed higher feature effects in that period. For longer prediction periods, NAVER search volumes were still found to constitute an important variable, although with a lower feature effect. This finding suggests that search term use should be considered to maintain the predictive performance of models.

Conclusions: NAVER search volumes were important variables in short- and long-term prediction, with higher feature effects for predicting new daily COVID-19 cases in the first 6 months of the outbreak. Similar results were also found for death predictions.

(J Med Internet Res 2021;23(12):e34178) doi:10.2196/34178

KEYWORDS

prediction; internet search; COVID-19; South Korea; infodemiology

Introduction

COVID-19 is a new disease entity that has caused a global pandemic, with more than 200 million cases and 4.5 million deaths since it was first reported at the end of December 2020 [1]. In contrast to the previous outbreaks of SARS and Middle East respiratory syndrome (MERS) that spread in clustered countries, COVID-19 exhibited massive disease transmission and longer periods of spread, even with implementation of multiple public health measures. Given this situation, predictions could greatly help in health resource management [2], particularly in terms of human resources and medical equipment deployment [3], as well as in preparing for upcoming future waves [4]. This approach will be beneficial for policy makers and health care managers [2], both in national government and at the level of local authorities [3].

However, as a new entity, COVID-19's disease dynamics seemed difficult to predict [5]. Most existing COVID-19 prediction models are highly dependent on confirmed cases, which may lag behind underlying infections [6]. The number of confirmed cases may only represent the number of people who have sought medical attention due to the occurrence of moderate to severe symptoms [5]. Therefore, external factors need to be included in models to increase the accuracy of these models.

One of the most common emerging external variables included in COVID-19 prediction models is comprised of internet search data. These data are collected during information-seeking activities on Google, NAVER, Daum, Baidu, and other search engines. Studies used information-seeking activities are part of infodemiology studies. The term "infodemiology" was first introduced by Eysenbach [7] in 2002 as an acronym of information epidemiology. This field aimed to analyze online information in terms of its distribution and determinants for public health–related purposes [8]. In addition, infodemiology is a fast-growing field of research that can be assessed both from demand- and supply-side studies [9]. Search engine query data in infodemiology studies are used in demand-based studies, which may have several advantages in the case of the COVID-19 pandemic. An increase in search data has commonly preceded traditional COVID-19 metrics [5,10,11]; as such, these data may provide a real-time indication of symptoms in a population [6]. Therefore, constructed models could possibly detect new waves or peaks at an earlier stage of the outbreak [5].

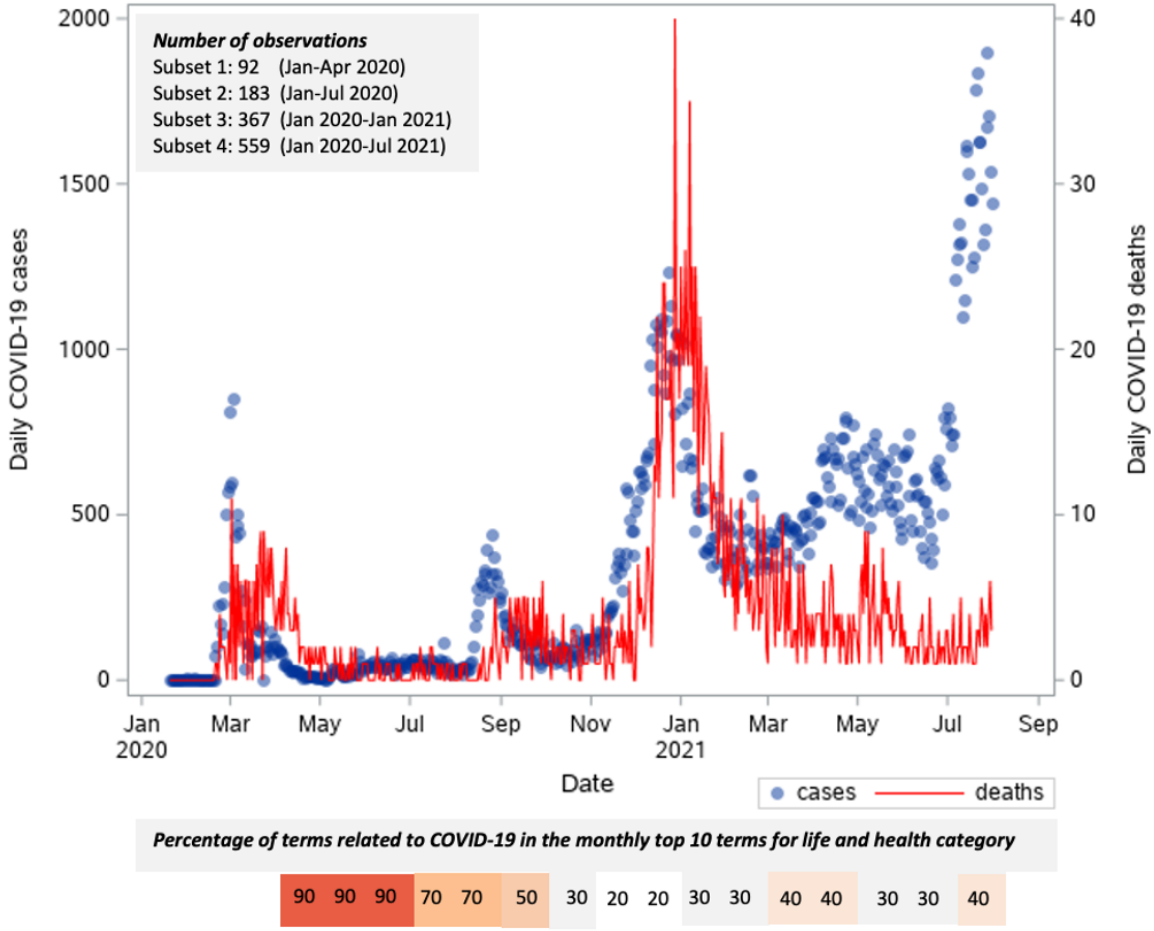
A study by Rabiolo et al [12] found that models that included search data performed better than those that did not include the search volumes in the first month of outbreak prediction. Similar findings were also shown in two previous analyses—one studying data from Iran [13] and another studying data from India, the United States, and the United Kingdom [14]—for periods of 1 month and 3 months, respectively, after the first case was detected. However, other studies conducted in the United States demonstrated low accuracy in model prediction [15] and variability in model performances among states and time periods [16]. Both of those studies were constructed using time series data of less than 2 months. Hence, it remains unclear whether models incorporating online search volumes will lead to better predictive performances for longer periods of prediction of new daily COVID-19 cases and deaths. In this study, we assessed the predictive performance of NAVER search volumes at different pandemic stages in South Korea. Data were aggregated into four subsets: 3, 6, 12, and 18 months after the first case was reported. In brief, this study aimed to analyze whether search engine query data constitute an important variable for inclusion in models for short- and long-term prediction of new daily COVID-19 cases and deaths.

Methods

Data Sets

In this study, we used country-level case-related data, NAVER search volumes, and mobility data from Google and Apple. NAVER search volumes were retrieved from NAVER's website [17] using terms related to COVID-19 and popular terms as of July 31, 2021. Terms in Korean, followed by their English translation, included the following: 코로 나 바이러스 (coronavirus), 코로나 바이러스 테스트 (coronavirus test), 메르 스 (MERS), 마스크 (face mask), 사회적 거리두기 (social distancing), 신천지 (Shincheonji), kf94 마스크 (kf94 mask), 일회용 마스크 (disposable mask), 온도계 (thermometer), 손 소독제 (hand sanitizer), 마스크스트랩 (mask strap), and Kf80 마스크 (kf80 mask). NAVER search volumes were queried in the Korean language, with quotation marks used for terms with more than two words, for all types of searches, genders, and age groups. Mobility data were collected from Google's Community Mobility Reports [18] and Apple's Mobility Trends Reports [19]. In addition to case-based data, daily cumulative COVID-19 cases and deaths were downloaded from the country-level time series data repository from the Center for Systems Science and Engineering at Johns Hopkins University [20]. A detailed description of all data used in this study is given in Table 1. Case-related data were retrieved from January 20, 2020—when the first COVID-19 case was reported in South Korea—to July 31, 2021. NAVER search volumes and mobility data were queried with a lag of 3 days to include more-recent observations in the analysis. Data were then aggregated into four subsets: 3, 6, 12, and 18 months after the first case was reported (Figure 1). Moreover, we also retrieved the monthly top 10 terms in the life and health category from NAVER beginning in April 2020 (Multimedia Appendix 1).

Figure 1. Time series of new daily COVID-19 cases and deaths in South Korea from January 20, 2020, to July 31, 2021. The information at the bottom of the figure describes the percentage of terms related to COVID-19 per month, from April 2020 to July 2021, out of the monthly top 10 terms for the life and health category (N=10). The list of terms is provided in Multimedia Appendix 1.



Statistical Analysis

Explanatory variables (Table 1 [21]) were used to develop models for predicting new daily COVID-19 cases and deaths. The first 80% of the data in all subsets were used as the training set, and the remaining data served as the testing set. In order to determine the best-fitting model in each subset, generalized linear models (GLMs) with three different distributions (ie, normal, Poisson, and negative binomial) were developed, along with linear regressions (LRs) with lasso, adaptive lasso, and elastic net regularization.

Table 1. Data set description.

Data set[a]	Data description	Use
Case-based data	Daily cumulative cases and deaths; used to calculate new daily cases and deaths	Time series graph, correlation, and prediction analysis
Google Community Mobility data	Daily changes in time spent in six categorized places—retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas—compared to baseline days; median value from January 3 to February 6, 2020	Correlation and prediction analysis
Apple Mobility Trends data	Daily relative volume of direction requests, in driving and walking situations, in Apple Maps compared to a baseline volume on January 13, 2020	Correlation and prediction analysis
NAVER search volumes	Daily online searches made through NAVER search engines; data ranged from 0 to 100; queries were made based on 12 terms used in our previous study [21] and popular terms related to COVID-19 as of July 31, 2021, from the life and health category; data were retrieved using terms in the Korean language, with quotation marks used for terms of more than two words, for all types of searches, genders, and age groups	Correlation and prediction analysis

[a]All data sets include country-level data.

All analyses and visualizations were conducted using SAS Studio, which is part of SAS OnDemand for Academics (SAS Institute Inc). For the GLMs, proc hpgenselect in SAS was used to develop and test the model performance with stepwise selection and an α level of .05 in selecting variables for the model. Only statistically significant variables (P<.05) were

included in the model. Furthermore, proc glmselect in SAS was used to construct LR models with steps of 25 and with the lowest Akaike information criterion (AIC) values in defining model selection. The 25 model construction steps were chosen in order to provide sufficient steps to define the best model with the lowest AIC value. Root mean square error (RMSE) values were defined as loss functions to assess the performance of models in the four subsets..

Results

Characteristics of New Daily COVID-19 Cases, Deaths, Mobility, and Search Data

The first case of COVID-19 in South Korea was reported on January 20, 2020, as shown in Figure 1. In the first 3 months of the outbreak, the mean number of new daily cases was 116.02. During this period, massive numbers of coronavirus tests were conducted along with the strict implementation of the social distancing policy. On February 7, 2020, the first coronavirus test kit was approved [22], and the first coronavirus drive-through test center was opened on February 23, 2020 [23]. The curve of cases was flattened, which led to an easing of social distancing rules at the national level beginning on May 6, 2020. A contact tracing system called KI-Pass was also introduced during this period [24]. Thus, with implementation of strict public health measures, the average new daily cases in the first 6 months of the outbreak dropped to 75.50, which was lower than that in the first 3 months.

However, a surge of cases occurred in mid-August, which led to a reinstating of level 2 restrictions beginning on August 28, 2020, in conjunction with mandatory mask-wearing. On October 12, 2020, restrictions were eased throughout most of the country, although a huge surge of cases developed as of mid-November. Level 2 restrictions were then tightened again [24]. This wave of cases remained high until the early months of 2021.

The first COVID-19 vaccine in South Korea was rolled out on February 28, 2021 [25]. Through the end of May, more than 700,000 people were newly vaccinated each day, but this number began to decrease by the end of June [26]. In early July, only around 1665 people were being vaccinated each day [27]. During this period, an immense wave of cases occurred that led to implementation of level 4 social distancing rules for the greater Seoul area beginning on July 26, 2021 [28]. The number of cases during this wave was larger than that of the other waves since the first reported COVID-19 case in South Korea. Time series analyses of new daily COVID-19 cases showed that implementation of public health measures heavily impacted the progression of cases. The number of new daily deaths seemed to follow the dynamics of COVID-19 cases, which were relatively higher in the third wave and lower in the fourth wave.

During the four waves of COVID-19 cases in South Korea, searches using various terms related to COVID-19 were captured in the NAVER database. In Figure 1, percentages of terms related to COVID-19 in the life and health category are presented. Due to the limitations of data querying in NAVER, retrospective top searches are only displayed starting from April 2020. A list of the top monthly terms is provided in Multimedia Appendix 1.

Figure 1 shows that a high percentage (9/10, 90%) of COVID-19–related terms were used in searches until June 2020. Afterward, the percentages decreased in the remaining months, with the lowest percentage (2/10, 20%) in November and December 2020. Relatively constant percentages of 30% (3/10) to 40% (4/10) of COVID-19–related terms used in internet searches were found in 2021. These findings reveal a massive use of COVID-19–related terms during online information-seeking activities in the early phase of the outbreak and a decreased pattern during the longer periods of the outbreak. In addition, top searches were mostly related to face masks, along with thermometers in April 2020 and hand sanitizers in August and September 2020.

Furthermore, decreased trends of mobility captured by Google were found to resemble the dynamics of cases and deaths (Figures 2 and 3). This differed from Apple mobility data, which seemed to be higher in the first and second waves and increased as the fourth wave developed. Moreover, increased numbers of searches seemed to precede the surge in cases and deaths.

Figure 2. Time series of new daily COVID-19 cases, mobility data (top plots), and NAVER searches (bottom plots) in South Korea from January 20, 2020, to July 31, 2021.

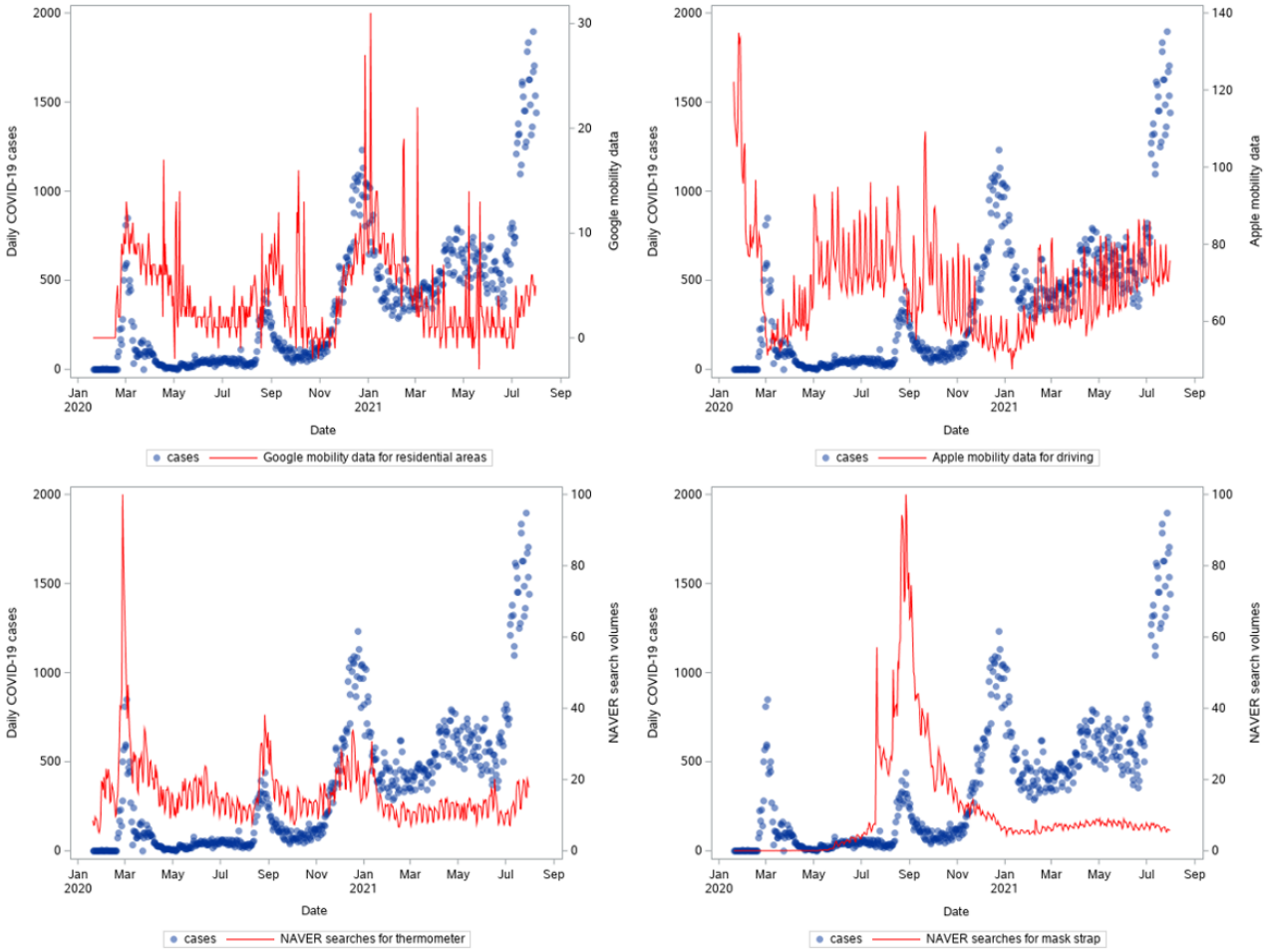
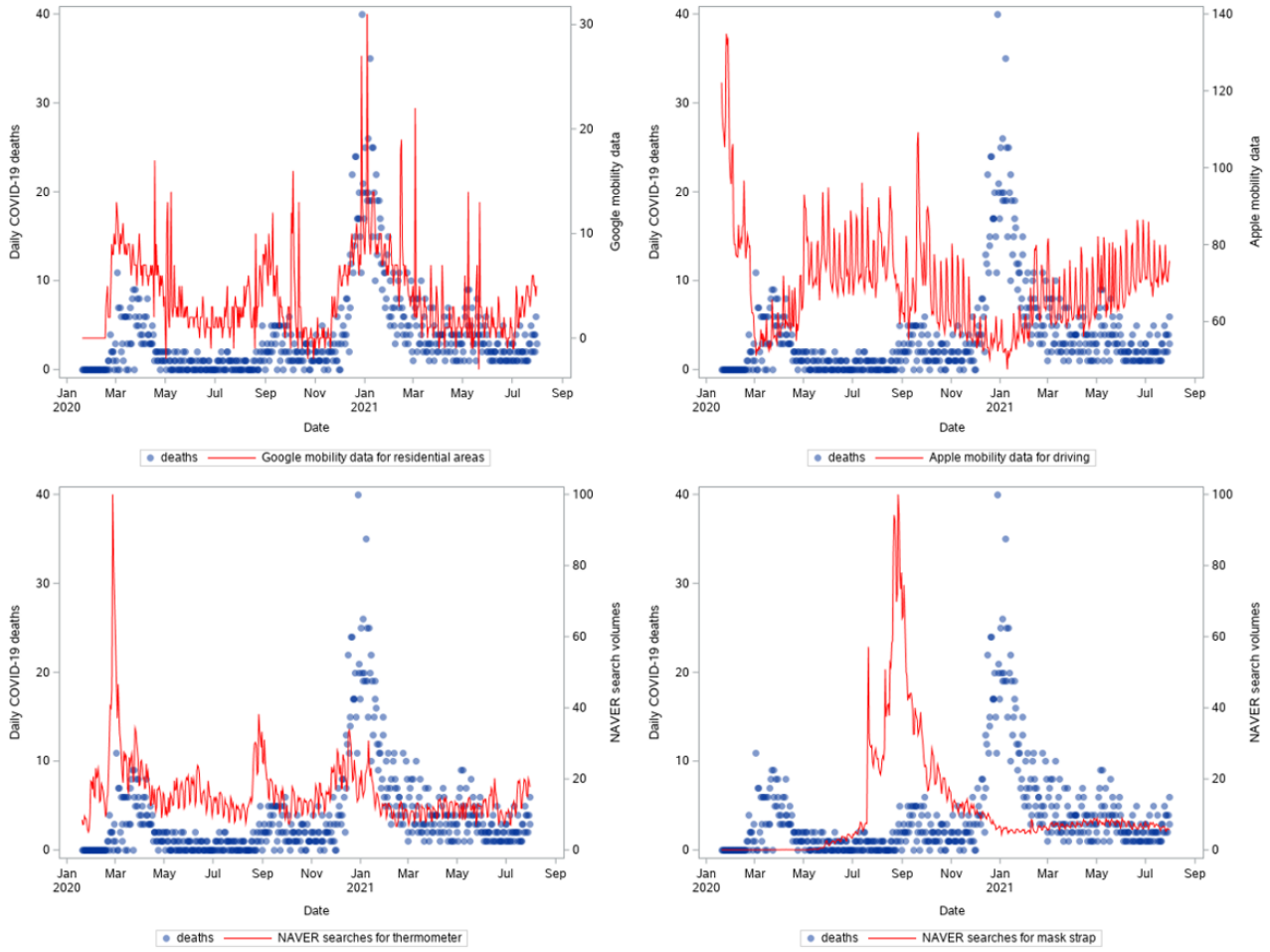


Figure 3. Time series of new daily COVID-19 deaths, mobility data (top plots), and NAVER searches (bottom plots) in South Korea from January 20, 2020, to July 31, 2021.



Correlations of New Daily COVID-19 Cases and Deaths With Explanatory Variables in the Training Sets

In the early stages of the outbreak, regarding subsets 1 and 2, new daily cases in the last 3 days (r=0.75, r=0.83), Google mobility data (retail and recreation: r=–0.82, –0.72; transit stations: r=–0.79, r=–0.70; residential areas: r=0.80), Apple mobility data (driving: r=–0.73; walking: r=–0.72), and NAVER search volumes (face mask: r=0.75; Shincheonji: r=0.83; thermometer: r=0.83, r=0.70) were highly correlated with new daily COVID-19 cases (Multimedia Appendix 2). In the third and fourth subsets, high correlations were only found between new daily COVID-19 cases and new daily cases in the last 3 days (r= 0.85, r=0.93). Moreover, moderate correlations were found between new daily COVID-19 cases in the third subset and Google mobility data (retail and recreation: r=–0.53) and between new daily COVID-19 cases in the last subset and new daily deaths in the last 3 days (r=0.62), Apple mobility data (driving: r=–0.62), and NAVER search volumes (disposable mask: r=–0.55). Negative correlations were mostly found between new daily COVID-19 cases and mobility data, which showed a decrease in the public's mobility during the pandemic period, particularly in the early stage of the outbreak. Negative correlations were also observed between new daily COVID-19 cases in the last subset of NAVER search volumes.

For new daily COVID-19 deaths, high correlations were only found for Apple mobility data (driving: r=–0.72; walking: r=–0.73) and NAVER search volumes (social distancing: r=0.72) in the first subset, and new daily cases in the last 3 days (r=0.71) and deaths (r=0.72) in the last subset. Similar to results in new daily COVID-19 cases, most of the negative correlations were found in mobility data in all subsets and NAVER search volumes in the last subset. Only Google mobility data for residential areas were positively correlated in all sets with both new daily COVID-19 cases and deaths. Results of the correlation analysis showed that higher correlations tended to be found in case-based data as the outbreak progressed, while reverse findings were found in mobility and internet search data.

Model Performance

GLMs with a Poisson distribution performed better as compared to the other models in predicting new daily COVID-19 cases in the first subset (Table 2). This finding suggests that at the early stage of the COVID-19 outbreak in South Korea, new daily cases more closely resembled a Poisson distribution. Later, in the second subset, the distribution of cases tended to be normally distributed, leading to a GLM with a normal function becoming the best performing model. GLMs with Poisson and negative binomial distributions resulted in larger RMSE values, which suggest that the distribution of cases in this subset did not follow those distributions that tended to be skewed.

In the third and fourth subsets, the LR without regularization (GLM1) and the LR with regularization (LR1-3) performed very similarly (Figure 4). This finding shows that GLMs performed better in the first 6 months of the outbreak. Over a longer period, LR models with regularization outperformed the GLMs. In addition, better performance of the model was found in predicting new daily deaths compared to new daily cases (Figures 4 and 5). For death predictions, the best performing models were the GLM with a negative binomial function in the first, second, and fourth subsets, and the LR with adaptive lasso regularization in the third subset.

Table 2. Assessment of the performance of the models.

Model	Subset 1[a], RMSE[b]		Subset 2[a], RMSE		Subset 3[a], RMSE		Subset 4[a], RMSE	
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
Predictions of new daily COVID-19 cases								
GLM1[c]	62.22	66.92	53.04	32.70[d]	48.01	378.94	85.75	219.22
GLM2[e]	43.71	29.29[d]	36.80	569,037.92	48.19	495.88	120.76	429.51
GLM3[f]	982.42	587.65	329.49	8,247,155.77	184.59	543.20	330.15	4161.61
LR1[g]	58.57	60.17	50.90	44.92	48.20	373.58	85.09	216.22[d]
LR2[h]	56.88	79.57	49.41	78.32	48.00	366.19[d]	84.52	216.70
LR3[i]	56.51	69.13	50.90	44.92	48.20	373.58	84.42	217.81
Predictions of new daily COVID-19 deaths								
GLM1	3.10	4.89	2.52	1.04	2.08	6.79	2.80	4.89
GLM2	3.24	5.52	2.71	0.47	2.23	7.65	2.82	5.26
GLM3	3.25	3.79[d]	2.72	0.19[d]	2.24	17.02	3.81	4.64[d]
LR1	3.05	4.95	2.62	1.71	2.16	5.21	2.75	5.23
LR2	3.04	4.50	2.61	0.70	2.19	4.82[d]	2.75	5.38
LR3	3.05	4.95	2.62	1.71	2.16	5.23	2.75	5.23

[a]Subsets 1 to 4: 3, 6, 12, and 18 months after the first case was reported in South Korea, respectively.

[b]RMSE: root mean square error.

[c]GLM1: generalized linear model with a normal distribution.

[d]The lowest RMSE value in the test subset.

[e]GLM2: generalized linear model with a Poisson distribution.

[f]GLM3: generalized linear model with a negative binomial distribution.

[g]LR1: linear regression model with lasso regularization.

[h]LR2: linear regression model with adaptive lasso regularization.

[i]LR3: linear regression model with elastic net regularization.

Figure 4. Time series of new daily COVID-19 cases in South Korea from January 20, 2020, to July 31, 2021, and predicted values in the generalized linear models (GLMs) and linear regression (LR) models. GLM1: GLM with a normal distribution; GLM2: GLM with a Poisson distribution; GLM3: GLM with a negative binomial distribution; LR1: LR model with lasso regularization; LR2: LR model with adaptive lasso regularization; LR3: LR model with elastic net regularization; RMSE: root mean square error.

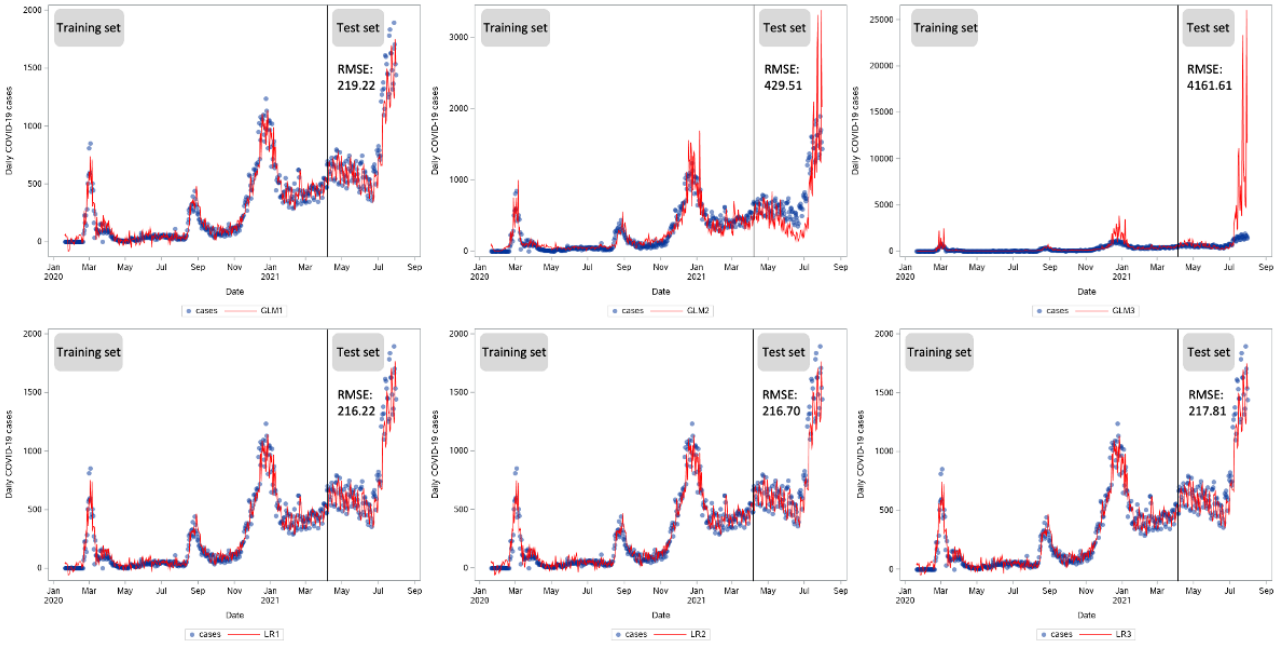
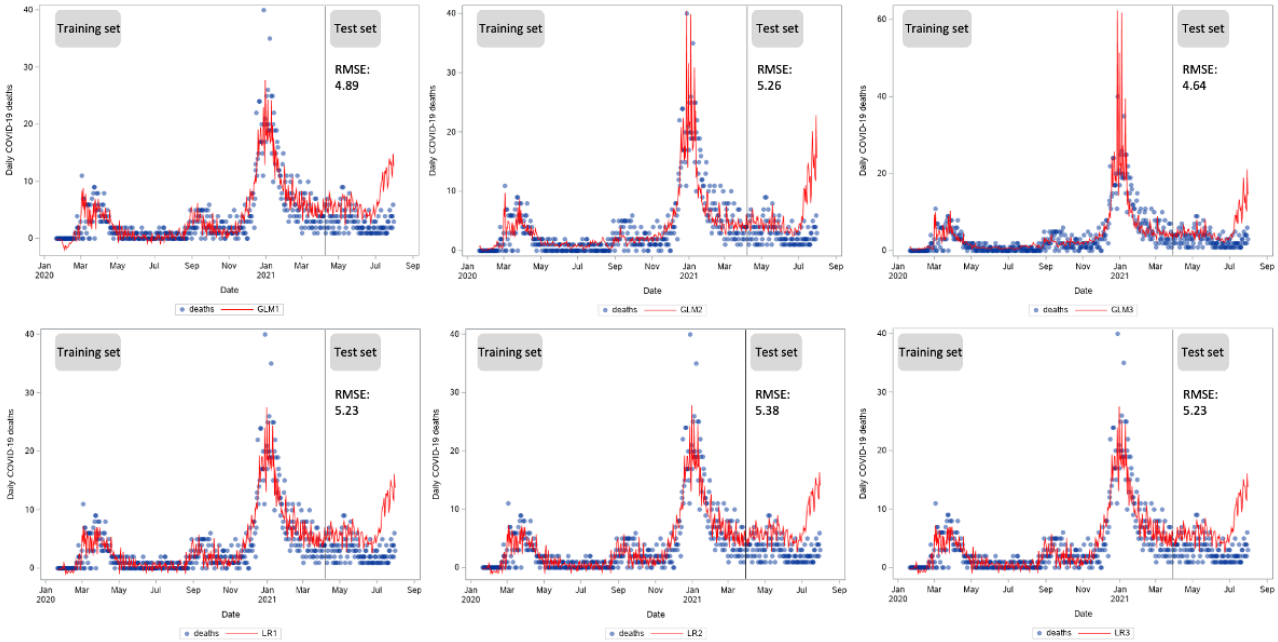


Figure 5. Time series of new daily COVID-19 deaths in South Korea from January 20, 2020, to July 31, 2021, and predicted values in the generalized linear models (GLMs) and linear regression (LR) models. GLM1: GLM with a normal distribution; GLM2: GLM with a Poisson distribution; GLM3: GLM with a negative binomial distribution; LR1: LR model with lasso regularization; LR2: LR model with adaptive lasso regularization; LR3: LR model with elastic net regularization; RMSE: root mean square error.



Feature Effects

In terms of the importance of the variables, NAVER search volumes produced higher parameter estimates in the models compared to case-based variables, Google mobility variables, and Apple mobility variables in the first and second subsets (Multimedia Appendix 3) in predicting new daily COVID-19 cases. This finding infers that NAVER search volumes might have affected the model performances to a greater extent and illustrates the usefulness of those variables, particularly searches for "thermometer" and "mask strap." However, parameter estimates of NAVER search volumes tended to have decreased in the third and fourth subsets. Higher parameter estimates were found in Google mobility data (ie, residential areas, transit stations, and workplaces) along with Apple mobility data (ie, driving) and case-based data (ie, new daily deaths in the last 3 days).

In this study, inclusion of NAVER searches for "thermometer" in models with longer periods seemed to be beneficial. In

addition, negative values of parameter estimates were found for most of the mobility data, except for the residential type, in all subsets. Negative parameter estimate values were also found in NAVER searches for "coronavirus," "coronavirus test," "MERS," "face mask," "kf80 mask," "disposable mask," "Shincheonji," and "hand sanitizer" in the third and fourth subsets.

In contrast to the abovementioned results, predictions of new daily deaths showed similar values of parameter estimates for case-based variables, mobility data, and NAVER search volumes (Multimedia Appendix 4). The numbers of variables included in the model were relatively fewer in the first and second sets compared with that in the third and fourth sets. However, most of the NAVER search volume variables were still included in the model, even in the last subset. Negative parameter estimate values were found for most of the mobility data, except workplace and residential types, in all subsets. Similarly, negative parameter estimate values for NAVER searches were also found in all sets. However, positive parameter estimate values were seen in all sets for case-based variables.

Discussion

Principal Findings

This study demonstrated an easy and explainable approach for determining the predictive performance of NAVER search volumes in four different subsets: 3, 6, 12, and 18 months after the first case was reported in South Korea. Subsets were used to create scenarios to analyze whether search engine query data are important variables for inclusion in models for short- and long-term prediction. In this study, we found that NAVER search volumes were useful variables in predicting new daily COVID-19 cases and deaths, particularly in the first 6 months of the outbreak. For longer prediction periods, NAVER search volumes were still found to constitute an important variable, although with a lower feature effect. In addition, this study discussed the role of search engine query data in infodemiology studies during the COVID-19 pandemic.

Short- and Long-Term Predictive Performances

Findings exhibited massive use of COVID-19–related terms for information-seeking activities at the early stage of the outbreak, which decreased over the longer period of the outbreak. This indicated a huge surge in information searches in the early months of the outbreak, as only limited COVID-19–related information was circulating. However, in later periods, extensive amounts of information were available, such as online news and reports by health experts [29]. Thus, these induced decreases in online information-seeking practices, which were observed from search term use. Beginning in April 2020, top searches were mostly related to face masks (Multimedia Appendix 1). A previous study in South Korea [21] showed increased in searches for various keywords concerning national and international events in the first 2 months of the pandemic. Similar results were also found in a worldwide setting [10], Taiwan [30], the Philippines [31], and the United States [32,33]. In addition, changes in the use of terms might indicate public concerns throughout the pandemic stages. In the case of South Korea, searches tended to be more related to logistical needs, including face masks, thermometers, and hand sanitizers for certain months in the longer period of the pandemic.

In terms of the correlation analyses, negative correlations were observed in the last subset for NAVER search volumes, which demonstrated a decline of searches as the number of cases increased. This finding is in line with an earlier study [16]. Moreover, lower correlation coefficients were found in search data as the outbreak progressed. This indicated the public's concern in terms of online information searches related to the ongoing outbreak, which tended to change over time. In addition to the prediction models, GLMs with different types of distribution functions may have been beneficial in predicting new daily COVID-19 cases and deaths in the early stages of the outbreak. Nonnormal distributions of cases and deaths could be better predicted using a Poisson or negative binomial function. Over a longer period, as the distribution of cases and deaths changed more into a normal distribution, LR models with regularization may have outperformed the GLMs. The use of regularization could also be important in preventing overfitting due to increased numbers of possible terms used in the longer period of prediction. This study also found that better performances of models were achieved in predicting new daily deaths compared to new daily cases, as found in a previous study [34]. This finding suggests higher variability of time series components (ie, trend, seasonality, and error) in new daily COVID-19 cases, which affected the prediction performances.

Furthermore, feature effects in the models showed that NAVER search volumes were useful variables in predicting new daily COVID-19 cases, particularly in the first 6 months of the outbreak. Searches related to logistical needs, particularly for "thermometer" and "mask strap," showed higher feature effects in that period. Compared to previous studies [12-14], terms with higher feature effects in the models were varied, from COVID-19–related terms, symptoms, and preventive measures. For longer prediction periods, NAVER search volumes were still found to be important variables, although with lower feature effects demonstrated from values of the parameter estimates. This result suggests that term use should be considered to maintain the prediction performance. This task may be subject to several challenges, since terms selected from top searches might not always perform as important variables in the model. Therefore, extensive keyword queries are needed to ensure that all possible and related terms are included in the model development.

Lastly, NAVER search volumes were also found to be beneficial in predicting new daily COVID-19 deaths, even for longer periods. Negative parameter estimate values for NAVER searches in the models were in line with results of the correlation analyses. This possibly suggests a decline in searches as the number of cases increases, although NAVER search engine query data were still regarded as useful variables for inclusion in the models.

The Role of Internet Search Data in Infodemiology Studies of COVID-19

As the COVID-19 pandemic has emerged, infodemiological studies related to COVID-19 grew exponentially. In general, such studies can be divided into three major subjects: studies

to understand community online search behaviors, preliminary studies to assess possible use of search data for prediction purposes, and prediction analyses. Studies to understand community online search behaviors are mostly aimed at assessing how the public responds in online information-seeking practices during a pandemic situation. Studies conducted by Strzelecki [11], Effenberger et al [10], Springer et al [35], Husain et al [32], and Hu et al [36] are examples of this kind of study. Those studies used search engine query data to understand patterns of information-seeking behaviors, particularly in interpreting public interest toward the ongoing pandemic. Some studies [31,33,37] have also been designed specifically to understand essential health information searched for by the public as cases increased. In addition, these types of studies were also used to assess health risk communication strategies [30] and health risk perceptions [21].

For preliminary studies to assess the possibility of using search data for prediction purposes, most studies found high correlations between COVID-19 cases and online search data [4-6]. Some also exhibited highly correlated patterns in the preceding days [2] and weeks [3]. Therefore, internet searches have become a potential data source for predicting COVID-19–related metrics. However, limited studies are available that assess predictive performances of search volume models. Studies [12-14] conducted in the early months of the pandemic showed that proposed models that included search data performed better than those that did not include the search volumes. However, studies in the United States demonstrated low accuracy in model prediction [15] and variability in model performance among states and time periods [16].

Accordingly, in this study, we assessed the predictive performances of models that incorporated online search volumes. Data were aggregated into four subsets: 3, 6, 12, and 18 months of time series data. We intended to analyze whether search engine query data are important variables for inclusion in models for short- and long-term prediction of new daily COVID-19 cases and deaths. Results demonstrated promising use of NAVER search volumes for prediction tasks with higher feature effects in the first 6 months of the outbreak. Thus, this study provides an overview of using search data for predictive purposes in the context of a pandemic situation.

Limitations

Analyses reported in this study only drew from perspectives of a demand-based infodemiological study. This means that this research examined the information-seeking behavior through search engine queries [9], which potentially reflect sudden changes in users' online behaviors toward the ongoing pandemic [38]. Future analyses may need to take into account the supply-side analysis, incorporate other search engines' data sets, as well as retrieve wider terms in order to capture wider infodemiological patterns in the population. In addition, other dynamic explanatory variables, such as health policy indices, may need to be included in the models to increase the model performance.

Conclusions

NAVER search volumes were important variables with higher feature effects for predicting new daily COVID-19 cases, particularly in the first 6 months of the outbreak in South Korea. For longer periods, NAVER search volumes were still found to be important variables, although search term use should be considered, as more specific terms need to be used. Similar results were also found for death predictions. Likewise, GLMs with different types of distribution functions may be beneficial for use in the early stages of an outbreak. In longer periods, LR models with regularization may outperform GLMs as the number of possible explanatory variables that can be used in the models increases.

Acknowledgments

This work was supported, via grants to ECYS, by the Ministry of Science and Technology in Taiwan (grants MOST109-2221-E-038-018 and MOST110-2628-E-038-001) and the Higher Education Sprout Project by the Ministry of Education in Taiwan (grant DP2-110-21121-01-A-13). This work was also supported, via a grant to ES, by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (grant 2021R1A6A1A10044154). The sponsors had no role in the research design or content of the manuscript for publication. The authors wish to acknowledge the Center for Systems Science and Engineering at Johns Hopkins University for use of their geographic information system dashboard and providing open-access data on daily cumulative COVID-19 cases and deaths in South Korea. In addition, the authors wish to acknowledge Google and Apple for allowing access to freely available data on community mobility, and NAVER for allowing access to the online search volumes.

Authors' Contributions

AH designed the study, performed the experiments, analyzed the data, and drafted and revised the manuscript. ES contributed analytical suggestions and revised the manuscript. AF made analytical suggestions. ECYS conceived the study, designed the experiments, and revised the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of monthly top terms in the life and health category from NAVER; terms have been translated into English.
[DOCX File , 21 KB - jmir_v23i12e34178_app1.docx]

Multimedia Appendix 2
Correlations of new daily COVID-19 cases and deaths with explanatory variables in the training sets.
[DOCX File , 25 KB - jmir_v23i12e34178_app2.docx]

Multimedia Appendix 3
Important variables included in the models for predicting new daily COVID-19 cases.
[DOCX File , 27 KB - jmir_v23i12e34178_app3.docx]

Multimedia Appendix 4
Important variables included in the models for predicting new daily COVID-19 deaths.
[DOCX File , 26 KB - jmir_v23i12e34178_app4.docx]

References

1. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. 2021. URL: https://covid19.who.int/ [accessed 2021-09-01]
2. 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):1-5 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
3. Ahmad I, Flanagan R, Staller K. Increased internet search interest for GI symptoms may predict COVID-19 cases in US hotspots. Clin Gastroenterol Hepatol 2020 Nov;18(12):2833-2834.e3 [FREE Full text] [doi: 10.1016/j.cgh.2020.06.058] [Medline: 32629121]
4. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 incidence using anosmia and other COVID-19 symptomatology: Preliminary analysis using Google and Twitter. Otolaryngol Head Neck Surg 2020 Sep;163(3):491-497 [FREE Full text] [doi: 10.1177/0194599820932128] [Medline: 32484425]
5. 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]
6. Ortiz-Martínez Y, Garcia-Robledo JE, Vásquez-Castañeda DL, Bonilla-Aldana DK, Rodriguez-Morales AJ. Can Google® Trends predict COVID-19 incidence and help preparedness? The situation in Colombia. Travel Med Infect Dis 2020;37:101703 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101703] [Medline: 32360323]
7. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
8. 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;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
9. Zeraatkar K, Ahmadi M. Trends of infodemiology studies: A scoping review. Health Info Libr J 2018 Jun;35(2):91-120. [doi: 10.1111/hir.12216] [Medline: 29729073]
10. 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]
11. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. Brain Behav Immun 2020 Aug;88:950-951 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.042] [Medline: 32311493]
12. Rabiolo A, Alladio E, Morales E, McNaught AI, Bandello F, Afifi AA, et al. Forecasting the COVID-19 epidemic by integrating symptom search behavior into predictive models: Infoveillance study. J Med Internet Res 2021 Aug 11;23(8):e28876 [FREE Full text] [doi: 10.2196/28876] [Medline: 34156966]
13. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Niakan Kalhori SR. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
14. Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PHJ. Forecasting spread of COVID-19 using Google Trends: A hybrid GWO-deep learning approach. Chaos Solitons Fractals 2021 Jan;142:110336 [FREE Full text] [doi: 10.1016/j.chaos.2020.110336] [Medline: 33110297]
15. 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]

16. Husnayain A, Chuang T, Fuad A, Su EC. High variability in model performance of Google relative search volumes in spatially clustered COVID-19 areas of the USA. Int J Infect Dis 2021 Aug;109:269-278 [FREE Full text] [doi: 10.1016/j.ijid.2021.07.031] [Medline: 34273513]
17. NAVER search volumes. NAVER. 2021. URL: https://datalab.naver.com/ [accessed 2021-11-06]
18. COVID-19 Community Mobility Reports. Google. 2021. URL: https://www.google.com/covid19/mobility/ [accessed 2021-11-06]
19. Mobility Trends Reports. Apple. 2021. URL: https://covid19.apple.com/mobility [accessed 2021-11-06]
20. 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]
21. Husnayain A, Shim E, Fuad A, Su EC. Understanding the community risk perceptions of the COVID-19 outbreak in South Korea: Infodemiology study. J Med Internet Res 2020 Sep 29;22(9):e19788 [FREE Full text] [doi: 10.2196/19788] [Medline: 32931446]
22. Normile D. Coronavirus cases have dropped sharply in South Korea. What's the secret to its success? Science. 2020 Mar 17. URL: https://www.science.org/news/2020/03/coronavirus-cases-have-dropped-sharply-south-korea-whats-secret-its-success# [accessed 2020-09-04]
23. Kwon KT, Ko JH, Shin H, Sung M, Kim JY. Drive-through screening center for COVID-19: A safe and efficient screening system against massive community outbreak. J Korean Med Sci 2020 Mar 23;35(11):e123 [FREE Full text] [doi: 10.3346/jkms.2020.35.e123] [Medline: 32193904]
24. Kim JH, Ah-Reum An J, Oh SJJ, Oh J, Lee JK. Emerging COVID-19 success story: South Korea learned the lessons of MERS. Our World in Data. 2021 Mar 05. URL: https://ourworldindata.org/covid-exemplar-south-korea [accessed 2021-09-04]
25. Dyer P. Policy and institutional responses to COVID-19: South Korea. Brookings. 2021 Jun 15. URL: https://www.brookings.edu/research/policy-and-institutional-responses-to-covid-19-south-korea/ [accessed 2021-09-04]
26. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. Nat Hum Behav 2021 Jul;5(7):947-953 [FREE Full text] [doi: 10.1038/s41562-021-01122-8] [Medline: 33972767]
27. Sang-Hun C. Want a vaccine reservation in South Korea? Try waiting 111 hours. The New York Times. 2021 Jul 28. URL: https://www.nytimes.com/2021/07/28/world/asia/south-korea-vaccine-covid.html [accessed 2021-09-04]
28. 'Living with COVID-19' scheme. George Mason University. 2021. URL: https://masonkorea.gmu.edu/corona/national-regulations-in-korea [accessed 2021-09-04]
29. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. Emerg Infect Dis 2009 May;15(5):689-695 [FREE Full text] [doi: 10.3201/eid1505.081114] [Medline: 19402953]
30. 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]
31. Galido A, Ecleo JJ, Husnayain A, Chia-Yu Su E. Exploring online search behavior for COVID-19 preventive measures: The Philippine case. PLoS One 2021;16(4):e0249810 [FREE Full text] [doi: 10.1371/journal.pone.0249810] [Medline: 33831076]
32. Husain I, Briggs B, Lefebvre C, Cline DM, Stopyra JP, O'Brien MC, et al. Fluctuation of public interest in COVID-19 in the United States: Retrospective analysis of Google Trends search data. JMIR Public Health Surveill 2020 Jul 17;6(3):e19969 [FREE Full text] [doi: 10.2196/19969] [Medline: 32501806]
33. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
34. Basu S, Campbell RH. Going by the numbers: Learning and modeling COVID-19 disease dynamics. Chaos Solitons Fractals 2020 Sep;138:110140 [FREE Full text] [doi: 10.1016/j.chaos.2020.110140] [Medline: 32834585]
35. Springer S, Menzel LM, Zieger M. Google Trends reveals: Focus of interest in the population is on treatment options rather than theories about COVID-19 animal origin. Brain Behav Immun 2020 Jul;87:134-135 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.005] [Medline: 32387509]
36. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. J Glob Health 2020 Jun;10(1):011003 [FREE Full text] [doi: 10.7189/jogh.10.011003] [Medline: 32373339]
37. Springer S, Menzel LM, Zieger M. Google Trends provides a tool to monitor population concerns and information needs during COVID-19 pandemic. Brain Behav Immun 2020 Jul;87:109-110 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.073] [Medline: 32360607]
38. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): Systematic review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]

Abbreviations

AIC: Akaike information criterion
GLM: generalized linear model
LR: linear regression
MERS: Middle East respiratory syndrome
RMSE: root mean square error

Edited by C Basch; submitted 09.10.21; peer-reviewed by A Mavragani, JY Wang; comments to author 31.10.21; revised version received 07.11.21; accepted 09.11.21; published 22.12.21.

Please cite as:
Husnayain A, Shim E, Fuad A, Su ECY
Predicting New Daily COVID-19 Cases and Deaths Using Search Engine Query Data in South Korea From 2020 to 2021: Infodemiology Study
J Med Internet Res 2021;23(12):e34178
URL: https://www.jmir.org/2021/12/e34178
doi:10.2196/34178
PMID:34762064

Original Paper

Automatic Recognition and Analysis of Balance Activity in Community-Dwelling Older Adults: Algorithm Validation

Yu-Cheng Hsu[1], PhD; Hailiang Wang[2], PhD; Yang Zhao[3], PhD; Frank Chen[4], PhD; Kwok-Leung Tsui[1,5], PhD

[1]School of Data Science, City University of Hong Kong, Kowloon, Hong Kong

[2]School of Design, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[3]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou, China

[4]Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

[5]Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Corresponding Author:
Yang Zhao, PhD
School of Public Health (Shenzhen)
Sun Yat-sen University
Room 111, Unit 1, Gezhi Garden 3#, No. 132, East Outer Ring Road
Guangzhou Higher Education Mega Center
Guangzhou, 510000
China
Phone: 86 020 83226383
Email: zhaoy393@mail.sysu.edu.cn

Abstract

Background: Clinical mobility and balance assessments identify older adults who have a high risk of falls in clinics. In the past two decades, sensors have been a popular supplement to mobility and balance assessment to provide quantitative information and a cost-effective solution in the community environment. Nonetheless, the current sensor-based balance assessment relies on manual observation or motion-specific features to identify motions of research interest.

Objective: The objective of this study was to develop an automatic motion data analytics framework using signal data collected from an inertial sensor for balance activity analysis in community-dwelling older adults.

Methods: In total, 59 community-dwelling older adults (19 males and 40 females; mean age = 81.86 years, SD 6.95 years) were recruited in this study. Data were collected using a body-worn inertial measurement unit (including an accelerometer and a gyroscope) at the L4 vertebra of each individual. After data preprocessing and motion detection via a convolutional long short-term memory (LSTM) neural network, a one-class support vector machine (SVM), linear discriminant analysis (LDA), and k-nearest neighborhood (k-NN) were adopted to classify high-risk individuals.

Results: The framework developed in this study yielded mean accuracies of 87%, 86%, and 89% in detecting sit-to-stand, turning 360°, and stand-to-sit motions, respectively. The balance assessment classification showed accuracies of 90%, 92%, and 86% in classifying abnormal sit-to-stand, turning 360°, and stand-to-sit motions, respectively, using Tinetti Performance Oriented Mobility Assessment-Balance (POMA-B) criteria by the one-class SVM and k-NN.

Conclusions: The sensor-based approach presented in this study provided a time-effective manner with less human efforts to identify and preprocess the inertial signal and thus enabled an efficient balance assessment tool for medical professionals. In the long run, the approach may offer a flexible solution to relieve the community's burden of continuous health monitoring.

(J Med Internet Res 2021;23(12):e30135) doi:10.2196/30135

KEYWORDS

fall risk; balance; activity recognition; automatic framework; community-dwelling elderly

Introduction

Falls prevail among the aging population, and led to more than $30 billion in direct medical costs in 2015 [1]. Around 55% of unintentional injury deaths among older adults in the United States are due to falls [2]. Falls pose a threat to the physical and psychological aspects of older adults' health [3,4]. It is critical

to identify older adults at the risk of falls and take interventions in advance [5].

Fall risk factors can be grouped into extrinsic (environmental) factors and intrinsic factors (age, health status, and other factors derived from the human). Outpatients and community dwellers usually suffer less from illnesses; consequently, intrinsic mobility and balance are the most discriminative intrinsic indicators of falls [6]. Mobility and balance assessment tools, such as the Balance Evaluation Systems Test (BESTest) [7], the Tinetti Performance Oriented Mobility Assessment (POMA) [8], and the Berg Balance Scale (BBS) [9], have been popular in community and outpatient settings to assess the mobility and balance aspects of individuals [6]. Along with the use of these instruments, a trained health care professional observes the participants' motion(s) as they complete a series of tasks (eg, sit-to-stand, turning 360°, and stand-to-sit) and scores their performance based on medical expertise.

Nevertheless, such assessment tools carry disadvantages that prevent older adults from undergoing frequent fall assessments [10]. First, many mobility and balance assessments, such as the BESTest [7], the Tinetti POMA [8], and the BBS [9], take from 15 to 35 min to complete [6,11], which is time consuming and burdensome to implement on a community-wide scale. Second, traditional assessments heavily rely on observations made by medical professionals [6,10]. Such resource-demanding assessments become unaffordable, which is a phenomenon commonly observed in Hong Kong [12,13]. A review [14] reported that over one-third of elderly services units fail to fill their physiotherapist and occupational therapist vacancies. In short, conducting mobility and balance assessment requires time, human resources, and financial availability, which further discourages older adults from frequently checking their fall risk.

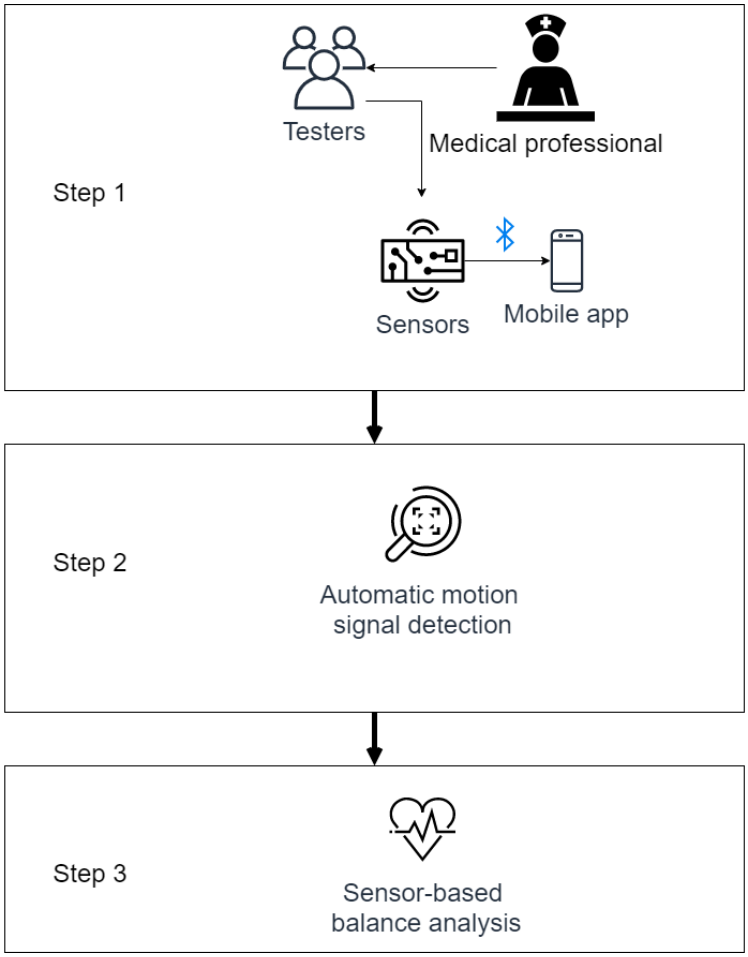
Thanks to the rapid development of information technology, the sensor provides a practical solution to this predicament nowadays. Commercial sensors (eg, accelerometer and gyroscope) are affordable to most community service sites or health care agents [15,16]. These sensors provide an objective measurement of motion that can support health professionals' decision making or act as a preliminary screening tool in the absence of professionals. Current research [17-19] has proven the utility of this approach by validating sensor-based assessments with clinical mobility and balance assessment tools. Researchers [20-25] have quantified and analyzed several sensor features that indicate motion and balance capability insufficiency. Halilaj et al [26] emphasized the importance of interpretability of sensor features and the model in this application for health care professionals and individuals. The majority have adopted statistical models to provide a statistical explanation of the work [10,16,27]. Some recent research [28-30] has argued that though the neural network approach might not be as interpretable as the previous research, it provides an accurate prediction. Sensor technology offers a quantitative method of studying human mobility and balance and coincides with clinical mobility and balance assessments.

Although there are advantages to the use of sensor technology, it is also important to acknowledge the residing problems. Ideally, individuals can perform these sensor-based assessments independently. In practice, irrelevant signals that come before and after performing the testing motion might be recorded. Therefore, an additional process to remove those irrelevant signals becomes necessary. Existing studies have relied on an additional research assistant's manual observations [28,31-33] or crafted features specific to motion [34] for identifying the motion signal, which will be assessed in the sensor-based assessments. In addition, although crafting features and building a heuristic algorithm [34,35] to detect motion provide a scientific solution, it is limited to specific, well-studied activity. This limitation makes detection hard to be generalized and transferred to different motion analyses. The additional manual efforts for identifying sensors or recording signals discourage health care professionals and community users from adopting sensor-based approaches [36].

Human activity recognition (HAR) is a practical remedy to relieve this burden. The majority of developed methods adopt machine learning models [37-39] or deep learning [40-45] to build on and validate renowned public data sets, such as the University of California, Irvine (UCI)-HAR [46] and UniMib SHAR [47], that target healthy adults aged 30 years or less. In contrast, only a few research works [48,49] focus on older adults. These studies [37-39] and reviews [43] claim that HAR techniques can be applied for health care purposes. Nonetheless, to the best of our knowledge, there is no existing work that explicitly delineates the combination of HAR and sensor-based mobility and balance assessments as completely automated assessments.

This study illustrates an automatic sensor-based framework (Figure 1) of motion and balance assessment. We hypothesized that the motion detection method is a solution for better automating sensor-based balance assessments. This framework resolves the requirement of human input in the preprocessing stage and completes the whole automation data pipeline. The developed framework aims to solve a simplified motion detection problem by leveraging the application scenario and a motion evaluation according to the Tinetti POMA-B grading standard. This study embodies this framework with deep-learning motion detection and sensor-based mobility and balance assessment. The major contributions of the proposed method are two-fold. First, few existing studies integrate motion detection and sensor-based balance assessment to form the data analysis pipeline, to the best of our knowledge. Second, the proposed method requires less human effort to identify and preprocess the inertial signal and enable a more efficient sensor-based balance assessment tool for medical professionals. This framework offers a flexible automatic solution to relieve the community's burden during large-scale implementation, such as long-term balance monitoring.

Figure 1. Overview of the developed framework.



Methods

Recruitment

In total, 59 community-dwelling elders (19 males and 40 females; mean age = 81.86 years, SD 6.95 years) participated in our study from September 2019 to December 2019. Participants with behavioral problems (eg, violence), unstable mental status (eg, paranoia), communication problems (eg, dialect), or severe hearing impairment were excluded from the study. Written informed consent was obtained from all participants before performing data collection. The imbalanced gender distribution was a direct reflection of the outnumbered male participants enrolling in community services [50]. Among the participants, six had experienced falls within the past year. A detailed breakdown of the demographics is shown in Table 1. This study was approved by the Research Ethics Committee of City University of Hong Kong (reference no. 3-2-201803-02).

Table 1. Demographics of all participants and group difference according to Tinetti grading items (N=59; 19 males and 40 females; mean age 81.86 years, SD 6.95 years).

Score	Age (years), mean (SD)	Males:females (n)
Sit-to-stand		
<4	86.64 (6.05)	6:8
4	80.37 (6.79)	13:32
Turning 360°		
<2	85.55 (5.88)	6:5
2	81.02 (6.90)	13:35
Stand-to-sit		
<2	87.00 (5.15)	5:9
2	80.26 (6.65)	14:31

The score followed the Tinetti POMA grading guideline [8]. The Tinetti POMA grading guideline is the mobility and balance assessment outcome used in this study. Age affected the Tinetti POMA performance, but gender did not. Age showed a

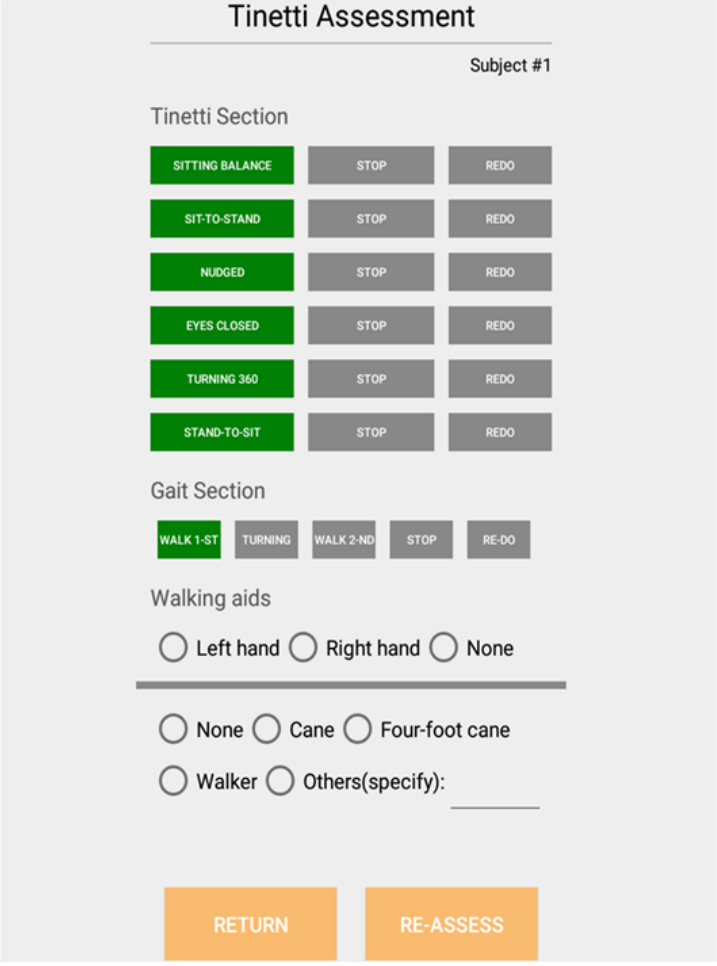
significant difference between the sit-to-stand score=4 and <4 groups (Student *t* test, *P*<.001). Age also showed a significant difference between the stand-to-sit score=2 and <2 groups (Student *t* test, *P*<.001). There was no gender difference in all of these tasks (Fisher exact test *P*=.99, 0.44, and 0.50 for sit-to-stand, turning 360°, and stand-to-sit motions, respectively.)

Experiment Protocol

In this step (Step 1 in Figure 1), balance assessments were conducted with an inertial sensor. Each participant was asked to perform three actions: standing up from an armless chair, turning around, and sitting down in the armless chair. There were 59 recorded inertial signals for each task from 59 participants. The three motions are commonly used in many functional assessments. All the three motions are adopted in the BBS [9], the Physical Performance Test (PPT) [50] and Tinetti POMA [8]; standing up and turning are adopted in the Fullerton Advanced Balance (FAB) [51] and the clinical Gait and Balance Scale (GABS) [52]; standing up and sitting down are adopted in the Postural Assessment Scale for Stroke Patients (PASS) [53], the Sensory Orientated Mobility Assessment Instrument (SOMAI) [54], and the Activity-based Balance Level Evaluation (ABLE) scale [55]. These three motions were considered sufficient to illustrate the idea of motion detection and evaluation in the developed framework.

A commercial inertial measurement unit (Wit-motion JY901B; including an accelerometer and a gyroscope at 3 axes with 16-bit resolution, sampling frequency 40 Hz, and a built-in Kalman filter) was attached on an elastic belt and placed on the L4 vertebra of each participant prior to each task. The inertial signal was transmitted via Bluetooth to the dedicated mobile application. The participants would then listen to the instruction and perform it accordingly. As soon as a participant initiated the task, a research assistant would press the button (green buttons and STOP buttons in Figure 2) on the app to mark the starting time and ending time, as shown in the clock on the mobile device. Figure 2 displays a screenshot of the data collection app.

Figure 2. Screenshot of the data collection app.



Before starting the assessment, each participant was required to perform a trial. Sit-to-stand and stand-to-sit motions were conducted using a straight-backed armless chair. Participants were asked to complete the turning task in a fast yet safe manner according to their judgment. These motions were conducted in line with Tinetti POMA guidelines. Participants were asked whether they needed a break for rest, but none needed it.

Automatic Motion Detection

In this part (Step 2 in Figure 1), we aimed to determine the starting and ending times for a known task from inertial signals containing only a known task (sit-to-stand, turning 360°, and stand-to-sit) and nontask activities. Since some participants needed extra time to comprehend and respond to the instruction, some recordings included pretask records. In addition, it also

took a few seconds for users to stop the recording after the participants completed the motion. Therefore, some sensor signals irrelevant to the motion might be included in the data. Since the motion to be analyzed was known in the prescribed mobility and balance assessment list, we only needed to detect the motion. The developed framework leverages these facts to reduce the complexity of motion detection to a simpler binary classification problem. We built separate classifiers for each task motion, allowing each piece of signal that contained a known task (ie, sit-to-stand, turning 360°, or stand-to-sit) to go through the data pipeline separately, as introduced below.

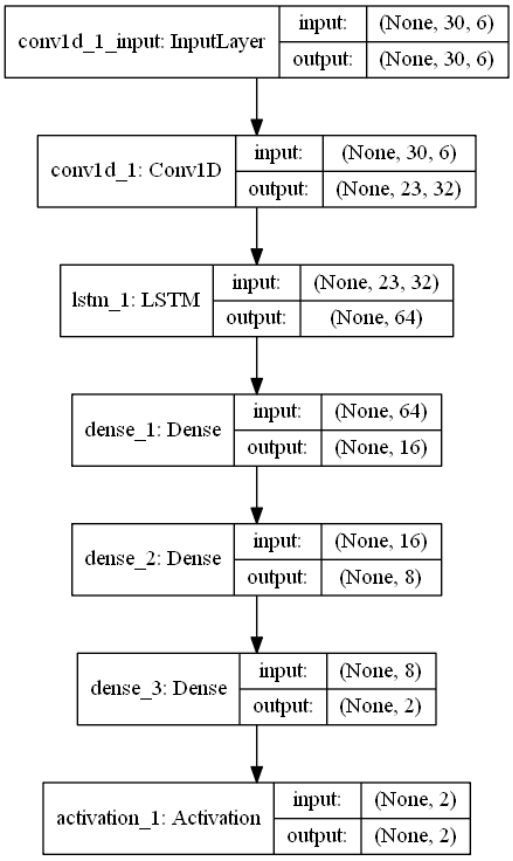
Preprocessing

The acceleration signal was detrended by subtracting the mean of the acceleration at each axis. The gyroscope signal was converted into the angular displacement from the starting point by integration. The gyroscope's drift issue was neglected as each task took less than 30 s to complete, and the drift rate is at 0.05°/s according to the product specification. Subsequently, each task signal was segmented from the preprocessed task signals into 0.75 s sliding windows with a step size of 0.03 s. The different sliding window sizes are examined and discussed in the Results and Discussion sections. The label of the sliding window segment is defined as the percentage of the class it covers. For example, a sliding window segment that contains 0.5 s of a sit-to-stand task and 0.25 s of the nontask motion would be labeled as 0.67 for the class sit-to-stand and 0.33 for the class nontask motion.

Convolutional LSTM Motion Detector

A convolutional long short-term memory (LSTM) neural network was built for detecting the starting and ending times for each task. The motion detection model takes the raw acceleration and angular displacement at three axes in the 0.75 sliding windows as input. The convolutional layers and LSTM layers can extract the temporal human dynamic inertial signals to construct the classification. The overall network structure is shown in Figure 3. This structure was inspired by a previous study [56], which used convolution networks to obtain the feature maps and LSTM structure to learn the temporal pattern.

Figure 3. Structure of a convolutional LSTM used in motion detection. LSTM, long short-term memory.

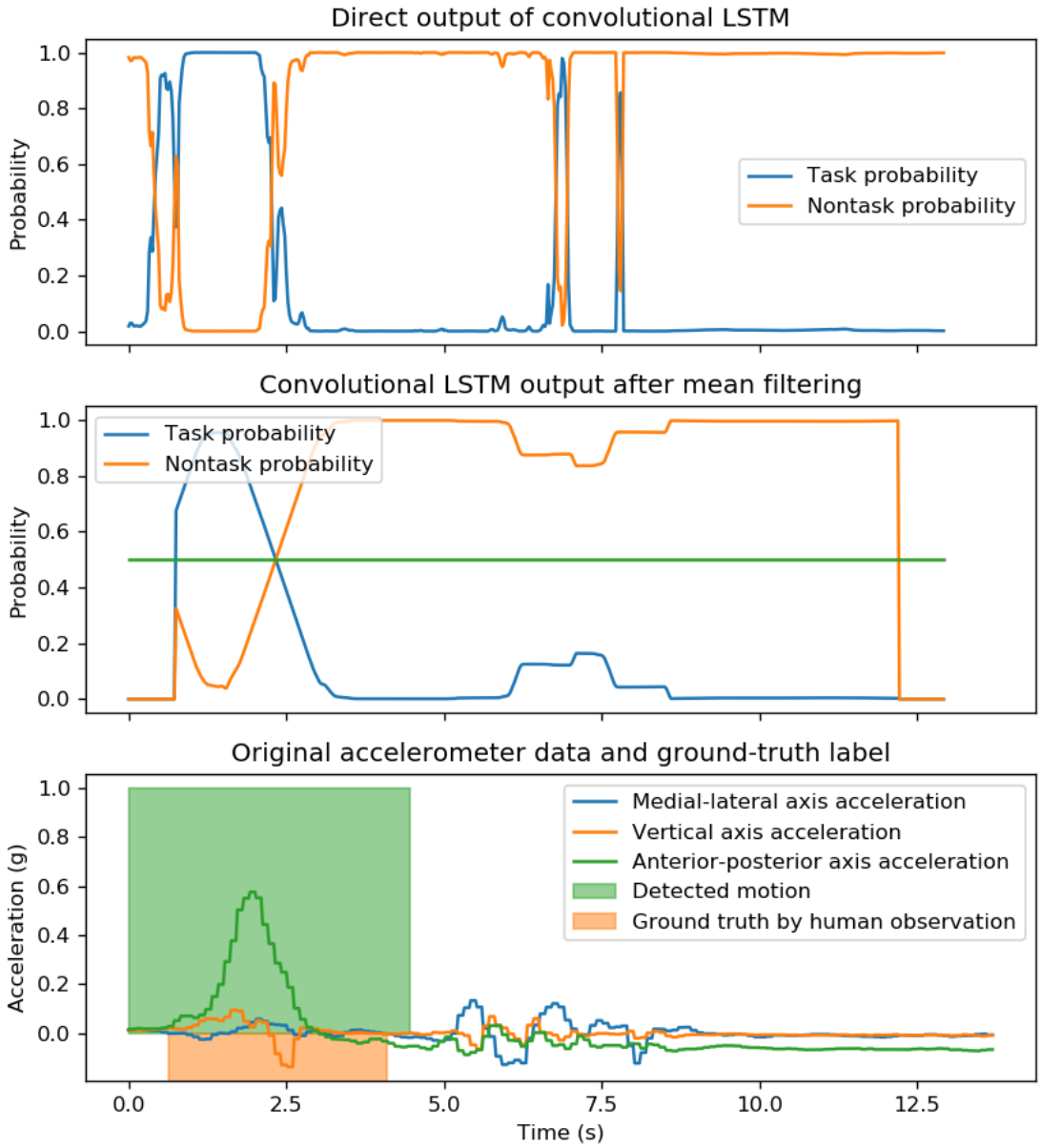


The convolution layer took the preprocessed sliding window segment as input and performed a convolution operation with 32 filters, kernel size 3. The LSTM layer is a specific implementation of the recurrent neural network that takes the previous state information into the current calculation. Finally, the densely connected layer operates by applying the activation function of the dot operation of weights. The first two dense layers did not apply any activation function, and the SoftMax function activated the last dense layer to calculate the percentage of labels containing and not containing the task.

Postprocessing and Labeling

The direct prediction of the convolutional LSTM neural network appeared to be noisy, as shown in the top image of Figure 4. This phenomenon was also observed in a similar network structure in HAR [56]. Therefore, the following postprocessing was performed to determine the beginning and end of the task.

Figure 4. Original signal and output of the convolution LSTM network before and after processing. LSTM: long short-term memory.



A mean filter was used to filter the convolutional LSTM neural network output with window sizes of 1, 1.25, and 1.5 s. These mean filter sizes used in this section are examined and discussed in the Results and Discussion sections. The output is shown in the middle image of Figure 4. Subsequently, the candidate intervals were identified according to the following two rules:

1. Beginning of the interval: The closest peak of nontask probability before the point where task probability starts to be greater than 0.5
2. End of the interval: The first peak of nontask probability after the point where task probability starts to be less than 0.5.

Peak detection was conducted using the SciPy package [57] with 0.5 thresholds.

Typically, only an interval would be identified. If more than one interval was identified, the following rules were applied accordingly to the different tasks. For sit-to-stand and stand-to-sit tasks, the time interval that contained the greatest anterior-posterior (AP) acceleration range was assigned to when the participant performed the motion. For the turning 360° task, intervals with cumulative angular movement from a starting time less than 360° along the turning axis (vertical axis) were selected. The final result is illustrated in the bottom image of Figure 4.

These rules were inspired from the biomechanical point of view for human motion [58], where AP acceleration drastically changes during the sit-to-stand or stand-to-sit motion. The assumption is that only one movement prevails in the signal recordings, as stated previously.

Sensor-Based Mobility and Balance Assessment

After the signal was automatically annotated, sensor-based mobility and balance assessment was performed to evaluate the corresponding motion (Step 3 in Figure 1). Sensor features that have been used in several previous studies were extracted from the detected signal and evaluated in relation to its Tinetti

POMA-B grading items. This part consisted of feature extraction and prediction modeling to achieve the goal.

Feature Engineering

Sit-to-Stand and Stand-to-Sit Tasks

The count of the acceleration peaks along the AP axis was extracted from the labeled data. Peak detection was conducted using the Scipy package [59] with the threshold 45% of the maximum value. A normal sit-to-stand transition usually shows one peak along the AP axis [58]. Multiple peaks imply the possibility that the motion was not smooth or multiple attempts were performed to achieve the task.

Turning 360° Task

The average turning speed along the vertical axis was extracted from the labeled data. Research [60] shows a significant difference in turning speed between fallers and nonfallers.

Prediction Modeling

Participants who did not receive full marks in the corresponding Tinetti POMA-B grading items were labeled as deviating from the norm in performing such tasks in this study. The goal of mobility and balance assessment is to identify older adults with insufficient mobility and balance capability and intervene as early as possible. Therefore, we aimed to identify older adults whose balance evaluation score was different from that of healthy adults (ie, full marks in the balance assessment). The sensor features introduced above were used to predict the corresponding Tinetti POMA-B grading items. A detailed description is presented in Table 2, and the corresponding sensor feature distribution between two populations is tabulated in Table 3.

Table 2. Tinetti POMA-B[a] task, grading items, and deviation from the healthy people criteria.

Task and grading item	Score	Deviation from healthy adults' criteria	Feature
Sit-to-stand		Tinetti POMA-B total sit-to-stand score <4	AP[b] acceleration peak count
Arises from the chair	~0-2		
Attempts to arise	~0-2		
Turning 360°		Tinetti POMA-B total turning 360° score <2	Average turning speed
Turns 360° continuously	~0-1		
Turns 360° steadily	~0-1		
Stand-to-sit		Tinetti POMA-B total stand-to-sit score <2	AP acceleration peak count
Sits down	~0-2		

[a]POMA-B: Performance Oriented Mobility Assessment-Balance.

[b]AP: anterior-posterior.

Table 3. Sensor feature distribution between normal and deviation from healthy participants.

Task	Sit-to-stand AP[a] acceleration peak count, mean (SD)	Turning 360° average turning speed (°/s), mean (SD)	Stand-to-sit AP acceleration peak count, mean (SD)
Healthy people	1.00 (0.29)	58.16 (19.22)	0.98 (0.15)
Deviating from healthy people	2.46 (2.02)	23.46 (13.96)	1.54 (0.75)

[a]AP: anterior-posterior.

A one-class support vector machine (SVM) [59], linear discriminant analysis (LDA), and k-nearest neighborhood (k-NN) from previous research works [31,61,62] were adopted in balance assessments.

Evaluation Metrics

Accuracy and the area under the curve (AUC) were used to evaluate the performance of this work. Accuracy was defined as the percentage of observations classified into the correct class, as (TP + TN)/(TP + TN + FP + FN), where TP is the true-positive class being classified as positive, TN is the true-negative class being classified as negative, FP is the false-negative class being classified as positive, and FN is the false-positive class being classified as negative. The AUC was obtained from the area under the receiver operating characteristic (ROC) curve. The ROC curve plots the sensitivity [TP/(TP + FN)]) against the false-positive rate [1 – (TN/{TN + FP})] at different levels of thresholds.

The agreement between the mobility and balance assessment results using the human-annotated label and the motion detection method was evaluated using the McNemar agreement test. The null hypothesis states that the two results show agreement in classifying the mobility and balance assessment evaluation results. The test statistic was calculated by $z^2 = (n_{12} - n_{21})^2/(n_{12} + n_{21})$, where n_{12} is the prediction result of when the manual label sensor feature shows positive but the motion detection label sensor feature shows negative. In contrast, n_{21} is the prediction result where the manual label sensor feature shows negative but the motion detection label sensor feature shows

positive. The test statistic followed a chi-square distribution with 1 degree of freedom.

Training and Testing Environment

Leave-one-subject-out (LOSO) cross-validation (CV) was used for training and testing. Each time, all the participants' inertial measurements except the i-th participant were used for training the motion detection and sensor-based balance assessment, and the inertial measurements from the i-th individual were used for testing. The convolutional LSTM neural network was built and trained using Keras [63] with TensorFlow [64].

The computer used for training was an Intel E5-2670 CPU with Nvidia Tesla K20 and CUDA version 10.2 on a Linux system. The batch size was set at 200, and the stopping rule was set for no improvement after 20 epochs. The rest of the computation was conducted using SciPy [57] and NumPy [65].

Results

Motion Detection

Our motion detection methods yielded moderate accuracy in detecting sit-to-stand, turning 360°, and stand-to-sit tasks from 85% to 88% at different levels (see Table 4). The difference in classification accuracy did not significantly vary between different mean filter sliding window sizes. The developed motion detection method detected the same motion performed by the participant not observed in the training stage, because the LOSO CV withheld the participant's motion signal as a testing set in each train and validation cycle. The training time and testing time showed that it took more than 1500 s to train the model, but detection was conducted within 0.3 s/participant.

Table 4. Accuracy of the automatic motion detection in different sliding windows.

Task	Accuracy (~Q1-Q3) mean filter 1 s sliding window (%)	Accuracy (~Q1-Q3) mean filter 1.25 s sliding window (%)	Accuracy (~Q1-Q3) mean filter 1.5 s sliding window (%)	Training time (s)	Testing time (s)
Sit-to-stand	87 (~82-95)	86 (~82-94)	85 (~78-93)	1681 (479)	0.21 (0.06)
Turning 360°	86 (~82-94)	86 (~79-95)	85 (~78-93)	2686 (495)	0.26 (0.09)
Stand-to-sit	88 (~85-94)	89 (~86-95)	88 (~83-94)	2186 (512)	0.24 (0.03)

Sensor-Based Mobility and Balance Assessment

The sit-to-stand task motion was well detected and classified using the developed method. The detector had 85%-87% accuracy in detecting sit-to-stand motion that was not previously observed in the training set. The extracted feature, the peak count along the AP axis, exhibited a strong ability to discriminate between normal and abnormal motion in the sit-to-stand task through the k-NN method at 90% accuracy. The high accuracy could be attributed to the fact that multiple attempts to rise generate multiple peaks in the signals [58], which aligns with the Tinetti POMA grading criterion sit-to-stand transition. LDA showed the least discriminative capability because of the Gaussian distribution assumption. The result shows that the developed motion detection method and balance evaluation model can predict abnormal sit-to-stand motion.

The turning 360° task motion was detected using the developed method at an accuracy ranging from 85% to 86%. The classification showed 92% agreement with the professional opinion using k-NN in terms of the Tinetti POMA turning motion outcome. These results indicate that Tinetti grading items may correlate with the turning speed of the participants. Previous studies [66,67] have also reported that the turning speed is correlated with certain clinical mobility and balance assessment tools.

The stand-to-sit task motion was detected at 88%-89% accuracy by the developed method. Compared with the other two tasks, however, sensor-based prediction showed the least accuracy in predicting the functional assessment result, at 86% accuracy using the one-class SVM. This may be ascribed to Tinetti POMA's grading to deduct a mark for participants who completed this task with the assistance of their arm, which occurred in the majority of the cases according to the report. This kind of information may not be revealed from the sensor signal, as the location of the sensor is at the participant's lower back. Consequently, it could result in weaker performance in classifying the category of older adults.

Discussion

Principal Findings

The sensor features from the manual label signal and the motion detection method yielded little difference, which barely affects the sensor-based balance assessment results. Both sit-to-stand AP peak counts and stand-to-sit peak counts showed no statistical difference (Wilcoxon signed-rank test P=.42 and .45, respectively). The average turn speed showed discrepancy (P=.03) with a mean difference of 5.68°/s. In summary, the motion detection label features showed statistically no difference in the sit-to-stand and stand-to-sit motions but a slight difference in turning.

Using the manual label motion signal features or the detected motion signal features showed no statistical difference in classifying the Tinetti POMA outcomes in all circumstances (Table 5). The McNemar test was conducted to determine whether there is a difference in the classification outcome between the manual label and the presented motion detection. The results revealed that there is no statistical difference. The motion detection accuracy was satisfactory enough to ensure that the sensor features were not affected by the detection method. Consequently, practitioners could be comfortable with adopting the motion detection method instead of traditional laborious works. None of the models rejected the null hypothesis

that both labels will yield the same result. This indicates that even though the motion detection method did not fully agree with the manual label, it still captures vital information in the sensor signal to predict mobility and balance assessments.

Table 5. Classification performance of sensor-based mobility and balance assessment using Tinetti-POMA-B[a] criteria.

Task and metrics	One-class SVM[b]	LDA[c]	k-NN[d]
Sit-to-stand			
AUC[e] (%)	84	62	82
Accuracy (%)	88	86	90
P value of McNemar test	>.99	>.99	>.99
Turning 360°			
AUC (%)	93	90	80
Accuracy (%)	68	86	92
P value of McNemar test	>.99	.25	>.99
Stand-to-sit			
AUC (%)	60	72	56
Accuracy (%)	86	83	80
P value of McNemar test	.5	.5	.125

[a]POMA-B: Performance Oriented Mobility Assessment-Balance.

[b]SVM: support vector machine.

[c]LDA: linear discriminant analysis.

[d]k-NN: k-nearest neighborhood.

[e]AUC: area under the curve.

Previous research [36] has reported clinician concerns about the real-time application of sensor-based balance assessments. The testing results in motion detection indicated only a little delay in obtaining the results of the developed framework. Therefore, the presented implementation showed acceptance by the clinician in terms of time efficiency. In contrast, a survey [68] reported that most of the Hong Kong community-dwelling older adults perceive motion-analyzing systems as useful. Accordingly, we believe that the framework would receive acceptance from both clinicians and older adults.

Limitations

There are four limitations of our study. First, we only analyzed sit-to-stand, turning 306°, and stand-to-sit motions rather than the complete mobility and balance assessment that covers comprehensive mobility and balance aspects. Nonetheless, motions used in this study have been frequently analyzed in several mobility and balance assessments to assess lower-body strength [69] and dynamic balance [70]. Other popular mobility and balance assessments, such as the Timed Up and Go (TUG) test [71], would be included in our future work. Second, instead of solving a HAR problem, this work leveraged the application scenario and broke it down into a detection problem, which only required an indication of the start and end times for the known task. This assumption helped the model to finesse the large variability in different motions in older adults. Third, we would like to expand the population and observations to enhance motion detection performance and balance assessments. In the framework discussed in this study, motion detection still relied on a certain amount of postprocessing, such as postprocessing after motion detection and feature engineering. This additional processing may become unnecessary with more observations for training. In addition, a greater sample size can also facilitate sensor-based balance evaluation with more complex modeling techniques and a one-off multiclass classifier to detect all the motions like in other works [28,30]. We can also observe that there are more healthy adults in the community than those with deviation from the healthy populations. A large sample size provides more in-depth insights into functional assessment with further complex feature engineering and feature selection methods. The last limitation is that the technology acceptance of this approach was not validated with the users. Though some related work with similar populations has shown good acceptance, as mentioned in the Discussion section, it is worth investigating how this approach can provide better user experience in the future plan.

Conclusions

We presented human motion detection in a large-scale health care assessment. Existing works [72,73] on HAR have achieved satisfactory results in detecting activities from nonmovement validation on public data sets. Nonetheless, most of the public data sets are limited in the subject count and focus on healthy subjects than normal life. The network structure of this study was also inspired by several previous works [56,74] on HAR. This research aimed to provide a sensor-based balance assessment approach to clinical decision support with less human effort. We illustrated a framework using an inertial sensor for balance assessment. The major contributions are two-fold. First, few existing studies have integrated motion detection and sensor-based balance assessment to form the data analysis pipeline, to the best of our knowledge. Second, our

method requires less human effort to identify and preprocess the inertial signal and enable a more efficient sensor-based balance assessment tool for medical professionals. This research examined the applicability of a deep-learning approach to detecting motion in the health care assessment context instead of general daily living.

This research illustrated the idea that functional assessment motions can be detected through HAR models. Therefore, the sensor data collection process can be conducted without additional labor if there is a sufficiently pretrained model in the future. Without the additional labor, the cost of sensor-based functional assessment can be reduced, providing more incentive to conduct large-scale implementation in identifying potential fallers in the community.

Acknowledgments

This work was partly supported by the CityU Provost Project (grant no. 9610406) and the National Natural Science Foundation of China (no. 71901188). The authors wish to thank Welfare Enterprises Association for its support to the study.

Authors' Contributions

YCH, HW, YZ, FC, and KLT were responsible for data collection. HW, YZ, FC, and KLT contributed to the study design and review of the manuscript. YCH conducted the analysis and writing of the report.

Conflicts of Interest

None declared.

References

1. Burns ER, Stevens JA, Lee R. The direct costs of fatal and non-fatal falls among older adults: United States. J Safety Res 2016;58:99-103 [FREE Full text] [doi: 10.1016/j.jsr.2016.05.001] [Medline: 27620939]
2. Kramarow E, Chen L, Hedegaard H, Warner M. Deaths from unintentional injury among adults aged 65 and over: United States, 2000-2013. NCHS Data Brief 2015:1-2.
3. Arfken CL, Lach HW, Birge SJ, Miller JP. The prevalence and correlates of fear of falling in elderly persons living in the community. Am J Public Health 1994;84(4):565-570. [doi: 10.2105/ajph.84.4.565] [Medline: 8154557]
4. Gillespie SM, Friedman SM. Fear of falling in new long-term care enrollees. J Am Med Dir Assoc 2007;8(5):307-313 [FREE Full text] [doi: 10.1016/j.jamda.2007.04.006] [Medline: 17570309]
5. Park S. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging Clin Exp Res 2018;30(1):1-16. [doi: 10.1007/s40520-017-0749-0] [Medline: 28374345]
6. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. J Gerontol A Biol Sci Med Sci 2001;56(12):M761-M766. [doi: 10.1093/gerona/56.12.m761] [Medline: 11723150]
7. Horak F, Wrisley D, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. Phys Ther 2009;89(5):484-498 [FREE Full text] [doi: 10.2522/ptj.20080071] [Medline: 19329772]
8. Tinetti ME, Franklin Williams T, Mayewski R. Fall risk index for elderly patients based on number of chronic disabilities. Am J Med 1986;80(3):429-434. [doi: 10.1016/0002-9343(86)90717-5]
9. Berg K. Balance and its measure in the elderly: a review. Physiother Can 1989;41(5):240-246. [doi: 10.3138/ptc.41.5.240]
10. Sun R, Sosnoff JJ. Novel sensing technology in fall risk assessment in older adults: a systematic review. BMC Geriatr 2018;18(1):14 [FREE Full text] [doi: 10.1186/s12877-018-0706-6] [Medline: 29338695]
11. Leddy A, Crowner B, Earhart G. Utility of the mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. J Neurol Phys Ther 2011:90-97. [doi: 10.1097/npt.0b013e31821a620c]
12. Siong K, Kwan MM, Lord SR, Lam AK, Tsang WW, Cheong AM. Fall risk in Chinese community-dwelling older adults: a physiological profile assessment study. Geriatr Gerontol Int 2016;16(2):259-265. [doi: 10.1111/ggi.12463] [Medline: 25655079]
13. Stevens JA, Phelan EA. Development of STEADI: a fall prevention resource for health care providers. Health Promot Pract 2013;14(5):706-714 [FREE Full text] [doi: 10.1177/1524839912463576] [Medline: 23159993]
14. Schoeb V. Healthcare service in Hong Kong and its challenges. China Perspect 2016;2016(4):51-58. [doi: 10.4000/chinaperspectives.7118]
15. Shany T, Redmond S, Marschollek M, Lovell N. Assessing fall risk using wearable sensors: a practical discussion. A review of the practicalities and challenges associated with the use of wearable sensors for quantification of fall risk in older people. Z Gerontol Geriatr 2012;45(8):694-706. [doi: 10.1007/s00391-012-0407-2] [Medline: 23184295]
16. Howcroft J, Kofman J, Lemaire ED. Review of fall risk assessment in geriatric populations using inertial sensors. J Neuroeng Rehabil 2013;10(1):91 [FREE Full text] [doi: 10.1186/1743-0003-10-91] [Medline: 23927446]
17. Senden R, Savelberg H, Grimm B, Heyligers I, Meijer K. Accelerometry-based gait analysis, an additional objective approach to screen subjects at risk for falling. Gait Posture 2012;36(2):296-300. [doi: 10.1016/j.gaitpost.2012.03.015] [Medline: 22512847]

18. Di Rosa M, Hausdorff JM, Stara V, Rossi L, Glynn L, Casey M, et al. Concurrent validation of an index to estimate fall risk in community dwelling seniors through a wireless sensor insole system: A pilot study. Gait Posture 2017;55:6-11. [doi: 10.1016/j.gaitpost.2017.03.037] [Medline: 28407507]
19. Greene BR, Doheny EP, Walsh C, Cunningham C, Crosby L, Kenny RA. Evaluation of falls risk in community-dwelling older adults using body-worn sensors. Gerontology 2012;58(5):472-480 [FREE Full text] [doi: 10.1159/000337259] [Medline: 22571883]
20. Giansanti D. Investigation of fall-risk using a wearable device with accelerometers and rate gyroscopes. Physiol Meas 2006;27(11):1081-1090. [doi: 10.1088/0967-3334/27/11/003] [Medline: 17028403]
21. Gietzelt M, Nemitz G, Wolf K, Meyer Zu Schwabedissen H, Haux R, Marschollek M. A clinical study to assess fall risk using a single waist accelerometer. Inform Health Soc Care 2009;34(4):181-188. [doi: 10.3109/17538150903356275] [Medline: 19919296]
22. Marschollek M, Nemitz G, Gietzelt M, Wolf K, Meyer Zu Schwabedissen H, Haux R. Predicting in-patient falls in a geriatric clinic: a clinical study combining assessment data and simple sensory gait measurements. Z Gerontol Geriatr 2009;42(4):317-321. [doi: 10.1007/s00391-009-0035-7] [Medline: 19543681]
23. Marschollek M, Rehwald A, Wolf KH, Gietzelt M, Nemitz G, Meyer zu Schwabedissen H, et al. Sensor-based Fall Risk Assessment: an Expert 'to go'. Methods Inf Med 2018;50(05):420-426. [doi: 10.3414/me10-01-0040]
24. Hua A, Quicksall Z, Di C, Motl R, LaCroix AZ, Schatz B, et al. Accelerometer-based predictive models of fall risk in older women: a pilot study. NPJ Digit Med 2018;1(1):25-28 [FREE Full text] [doi: 10.1038/s41746-018-0033-5] [Medline: 31304307]
25. Liu Y, Redmond SJ, Wang N, Blumenkron F, Narayanan MR, Lovell NH. Spectral analysis of accelerometry signals from a directed-routine for falls-risk estimation. IEEE Trans Biomed Eng 2011;58(8):2308-2315. [doi: 10.1109/tbme.2011.2151193]
26. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. J Biomech 2018;81:1-11 [FREE Full text] [doi: 10.1016/j.jbiomech.2018.09.009] [Medline: 30279002]
27. Galán-Mercant A, Cuesta-Vargas AI. Differences in trunk accelerometry between frail and nonfrail elderly persons in sit-to-stand and stand-to-sit transitions based on a mobile inertial sensor. JMIR Mhealth Uhealth 2013;1(2):e21 [FREE Full text] [doi: 10.2196/mhealth.2710] [Medline: 25098977]
28. Rivolta MW, Aktaruzzaman M, Rizzo G, Lafortuna CL, Ferrarin M, Bovi G, et al. Evaluation of the Tinetti score and fall risk assessment via accelerometry-based movement analysis. Artif Intell Med 2019;95:38-47. [doi: 10.1016/j.artmed.2018.08.005] [Medline: 30195985]
29. Howcroft J, Lemaire ED, Kofman J. Wearable-sensor-based classification models of faller status in older adults. PLoS One 2016;11(4):e0153240 [FREE Full text] [doi: 10.1371/journal.pone.0153240] [Medline: 27054878]
30. Buisseret F, Catinus L, Grenard R, Jojczyk L, Fievez D, Barvaux V, et al. Timed Up and Go and six-minute walking tests with wearable inertial sensor: one step further for the prediction of the risk of fall in elderly nursing home people. Sensors (Basel) 2020;20(11):3207 [FREE Full text] [doi: 10.3390/s20113207] [Medline: 32516995]
31. Simila H, Mantyjarvi J, Merilahti J, Lindholm M, Ermes M. Accelerometry-based Berg Balance Scale score estimation. IEEE J Biomed Health Inform 2014;18(4):1114-1121. [doi: 10.1109/jbhi.2013.2288940]
32. Ponti M, Bet P, Oliveira CL, Castro PC. Better than counting seconds: identifying fallers among healthy elderly using fusion of accelerometer features and dual-task Timed Up and Go. PLoS One 2017;12(4):e0175559 [FREE Full text] [doi: 10.1371/journal.pone.0175559] [Medline: 28448509]
33. Shahzad A, Ko S, Lee S, Lee J, Kim K. Quantitative assessment of balance impairment for fall-risk estimation using wearable triaxial accelerometer. IEEE Sensors J 2017;17(20):6743-6751. [doi: 10.1109/jsen.2017.2749446]
34. Ejupi A, Brodie M, Lord SR, Annegarn J, Redmond SJ, Delbaere K. Wavelet-based sit-to-stand detection and assessment of fall risk in older people using a wearable pendant device. IEEE Trans Biomed Eng 2017;64(7):1602-1607. [doi: 10.1109/tbme.2016.2614230]
35. Pham MH, Elshehabi M, Haertner L, Heger T, Hobert MA, Faber GS, et al. Algorithm for turning detection and analysis validated under home-like conditions in patients with Parkinson's disease and older adults using a 6 degree-of-freedom inertial measurement unit at the lower back. Front Neurol 2017;8:135 [FREE Full text] [doi: 10.3389/fneur.2017.00135] [Medline: 28443059]
36. Bergmann JHM, McGregor AH. Body-worn sensor design: what do patients and clinicians want? Ann Biomed Eng 2011;39(9):2299-2312. [doi: 10.1007/s10439-011-0339-9] [Medline: 21674260]
37. Rahim KNKA, Elamvazuthi I, Izhar L, Capi G. Classification of human daily activities using ensemble methods based on smartphone inertial sensors. Sensors (Basel) 2018;18(12):4132 [FREE Full text] [doi: 10.3390/s18124132] [Medline: 30486242]
38. Balli S, Sağbaş EA, Peker M. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. Meas Control 2018;52(1-2):37-45. [doi: 10.1177/0020294018813692]
39. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical human activity recognition using wearable sensors. Sensors (Basel) 2015;15(12):31314-31338 [FREE Full text] [doi: 10.3390/s151229858] [Medline: 26690450]

40. Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks. 2015 Presented at: ACM international conference on Multimedia; October 2015; Brisbane, Australia p. 1307-1310. [doi: 10.1145/2733373.2806333]
41. Ronao CA, Cho S. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst Appl 2016;59:235-244. [doi: 10.1016/j.eswa.2016.04.032]
42. Hassan MM, Uddin MZ, Mohamed A, Almogren A. A robust human activity recognition system using smartphone sensors and deep learning. Future Gener Comput Syst 2018;81:307-313. [doi: 10.1016/j.future.2017.11.029]
43. Ravi D, Wong C, Lo B, Yang G. Deep learning for human activity recognition: a resource efficient implementation on low-power devices. 2016 Presented at: 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN); 14-17 June 2016; San Francisco, CA p. 71-76. [doi: 10.1109/bsn.2016.7516235]
44. Chung S, Lim J, Noh KJ, Kim G, Jeong H. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. Sensors (Basel) 2019;19(7):1716 [FREE Full text] [doi: 10.3390/s19071716] [Medline: 30974845]
45. Niu X, Wang Z, Pan Z. Extreme learning machine-based deep model for human activity recognition with wearable sensors. Comput Sci Eng 2019;21(5):16-25. [doi: 10.1109/mcse.2018.110145933]
46. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz J. A public domain dataset for human activity recognition using smartphones. 2013 Presented at: European Symposium on Artificial Neural Networks; April 26, 2013; Bruges, Belgium p. 3. [doi: 10.1007/978-3-642-40728-4_54]
47. Micucci D, Mobilio M, Napoletano P. UniMiB SHAR: a dataset for human activity recognition using acceleration data from smartphones. Appl Sci 2017:1101. [doi: 10.20944/preprints201706.0033.v1]
48. Aicha AN, Englebienne G, van Schooten K, Pijnappels M, Kröse B. Deep learning to predict falls in older adults based on daily-life trunk accelerometry. Sensors (Basel) 2018;18(5):1654 [FREE Full text] [doi: 10.3390/s18051654] [Medline: 29786659]
49. Capela NA, Lemaire ED, Baddour N. Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. PLoS One 2015;10(4):e0124414 [FREE Full text] [doi: 10.1371/journal.pone.0124414] [Medline: 25885272]
50. Siu AL, Reuben DB. An objective measure of physical function of elderly outpatients: the Physical Performance Test. J Am Geriatr Soc 1990;38(10):1105-1112. [doi: 10.1111/j.1532-5415.1990.tb01373.x] [Medline: 2229864]
51. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil 2006;87(11):1478-1485. [doi: 10.1016/j.apmr.2006.07.263] [Medline: 17084123]
52. Thomas M, Jankovic J, Suteerawattananon M, Wankadia S, Caroline KS, Vuong KD, et al. Clinical gait and balance scale (GABS): validation and utilization. J Neurol Sci 2004;217(1):89-99. [doi: 10.1016/j.jns.2003.09.005] [Medline: 14675615]
53. Benaim C, Pérennou DA, Villy J, Rousseaux M, Pelissier JY. Validation of a standardized assessment of postural control in stroke patients: the Postural Assessment Scale for Stroke Patients (PASS). Stroke 1999;30(9):1862-1868. [doi: 10.1161/01.str.30.9.1862] [Medline: 10471437]
54. Tang P, Moore S, Woollacott MH. Correlation between two clinical balance measures in older adults: functional mobility and sensory organization test. J Gerontol A Biol Sci Med Sci 1998;53(2):M140-M146. [doi: 10.1093/gerona/53a.2.m140] [Medline: 9520921]
55. Ardolino E, Hutchinson K, Pinto Zipp G, Clark M, Harkema S. The ABLE scale: the development and psychometric properties of an outcome measure for the spinal cord injury population. Phys Ther 2012;92(8):1046-1054. [doi: 10.2522/ptj.20110257] [Medline: 22577065]
56. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors (Basel) 2016;16(1):115 [FREE Full text] [doi: 10.3390/s16010115] [Medline: 26797612]
57. Virtanen P, Gommers R, Oliphant T, Haberland M, Reddy T, Cournapeau D. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020:261-272. [doi: 10.1017/9781108778039.009]
58. Richards J. The Comprehensive Textbook of Clinical Biomechanics. Amsterdam, The Netherlands: Elsevier; 2018.
59. Tax DM, Duin RP. Support vector data description. Mach Learn 2004;54(1):45-66. [doi: 10.1023/b:mach.0000008084.60811.49]
60. Murphy M, Olson S, Protas E, Overby A. Screening for falls in community-dwelling elderly. J Aging Phys Act 2003:66-80. [doi: 10.1123/japa.11.1.66]
61. Polus JS, Bloomfield RA, Vasarhelyi EM, Lanting BA, Teeter MG. Machine learning predicts the fall risk of total hip arthroplasty patients based on wearable sensor instrumented performance tests. J Arthroplasty 2021;36(2):573-578. [doi: 10.1016/j.arth.2020.08.034] [Medline: 32928593]
62. Kargar B, Mollahosseini A, Struemph T, Pace W, Nielsen R, Mahoor M. Automatic measurement of physical mobility in Get-Up-and-Go Test using kinect sensor. 2014 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC; 2014; Chicago, IL. [doi: 10.1109/embc.2014.6944375]
63. Chollet F. Keras. 2015. URL: https://github.com/fchollet/keras [accessed 2021-11-06]
64. Martín A, Ashish A, Paul BE, Zhifeng C, Craig C, Greg S. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems Internet. URL: https://www.tensorflow.org/ [accessed 2020-10-01]

65. van der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 2011;13(2):22-30. [doi: 10.1109/mcse.2011.37]
66. Shubert TE, Schrodt LA, Mercer VS, Busby-Whitehead J, Giuliani CA. Are scores on balance screening tests associated with mobility in older adults? J Geriatr Phys Ther 2006;29(1):33-39. [doi: 10.1519/00139143-200604000-00007]
67. Gill T, Williams C, Tinetti M. Assessing risk for the onset of functional dependence among older adults: the role of physical performance. J Am Geriatr Soc 1995;43(6):603-609. [doi: 10.1111/j.1532-5415.1995.tb07192.x] [Medline: 7775716]
68. Lai CK, Chung JC, Leung NK, Wong JC, Mak DP. A survey of older Hong Kong people's perceptions of telecommunication technologies and telecare devices. J Telemed Telecare 2010;16(8):441-446. [doi: 10.1258/jtt.2010.090905] [Medline: 20870685]
69. Kirschke A, Kocur P, Wilk M, Dylewicz P. The Fullerton Fitness Test as an index of fitness in the elderly. Med Rehabil 2006:9-16.
70. Cheng F, Yang Y, Wang C, Wu Y, Cheng S, Wang H, et al. Factors influencing turning and its relationship with falls in individuals with Parkinson's disease. PLoS One 2014;9(4):e93572 [FREE Full text] [doi: 10.1371/journal.pone.0093572] [Medline: 24699675]
71. Shumway-Cook A, Brauer S, Woollacott M. Phys Ther 2000;80(9):896-903. [doi: 10.1093/ptj/80.9.896]
72. Sun J, Fu Y, Li S, He J, Xu C, Tan L. Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors. J Sensors 2018;2018:1-10. [doi: 10.1155/2018/8580959]
73. Akbari A, Jafari R. An autoencoder-based approach for recognizing null class in activities of daily living in-the-wild via wearable motion sensors. 2019 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; 2019; Brighton, UK. [doi: 10.1109/icassp.2019.8682161]
74. Zeng M, Nguyen L, Yu B, Mengshoel O, Zhu J, Wu P. Convolutional neural networks for human activity recognition using mobile sensors. 2014 Presented at: Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, MobiCASE; 2014; Austin, TX. [doi: 10.4108/icst.mobicase.2014.257786]

Abbreviations

ABLE: Activity-based Balance Level Evaluation
AP: anterior-posterior
AUC: area under the curve
BBS: Berg Balance Scale
BESTest: Balance Evaluation Systems Test
CV: cross-validation
FAB: Fullerton Advance Balance
FN: false negative
FP: false positive
GABS: Gait and Balance Scale
HAR: human activity recognition
k-NN: k-nearest neighborhood
LDA: linear discriminant analysis
LOSO: leave-one-subject-out
LSTM: long short-term memory
PASS: Postural Assessment Scale for Stroke Patients
POMA-B: Performance Oriented Mobility Assessment-Balance
PPT: Physical Performance Test
ROC: receiver operating characteristic
SOMAI: Sensory Orientated Mobility Assessment Instrument
SVM: support vector machine
TN: true negative
TP: true positive
TUG: Timed Up and Go

Edited by R Kukafka; submitted 03.05.21; peer-reviewed by J Peng, E Sadeghi-Demneh, H Li; comments to author 28.06.21; revised version received 23.08.21; accepted 22.09.21; published 20.12.21.

Please cite as:
Hsu YC, Wang H, Zhao Y, Chen F, Tsui KL
Automatic Recognition and Analysis of Balance Activity in Community-Dwelling Older Adults: Algorithm Validation
J Med Internet Res 2021;23(12):e30135
URL: https://www.jmir.org/2021/12/e30135
doi:10.2196/30135
PMID:34932008

Original Paper

Reliability of Commercial Voice Assistants' Responses to Health-Related Questions in Noncommunicable Disease Management: Factorial Experiment Assessing Response Rate and Source of Information

Caterina Bérubé[1], MSc; Zsolt Ferenc Kovacs[1], MSc; Elgar Fleisch[1,2,3], Prof Dr; Tobias Kowatsch[1,2,3], PhD

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

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

[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

Corresponding Author:
Caterina Bérubé, MSc
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
WEV G 214
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 633 8419
Email: berubec@ethz.ch

Abstract

Background: Noncommunicable diseases (NCDs) constitute a burden on public health. These are best controlled through self-management practices, such as self-information. Fostering patients' access to health-related information through efficient and accessible channels, such as commercial voice assistants (VAs), may support the patients' ability to make health-related decisions and manage their chronic conditions.

Objective: This study aims to evaluate the reliability of the most common VAs (ie, Amazon Alexa, Apple Siri, and Google Assistant) in responding to questions about management of the main NCD.

Methods: We generated health-related questions based on frequently asked questions from health organization, government, medical nonprofit, and other recognized health-related websites about conditions associated with Alzheimer's disease (AD), lung cancer (LCA), chronic obstructive pulmonary disease, diabetes mellitus (DM), cardiovascular disease, chronic kidney disease (CKD), and cerebrovascular accident (CVA). We then validated them with practicing medical specialists, selecting the 10 most frequent ones. Given the low average frequency of the AD-related questions, we excluded such questions. This resulted in a pool of 60 questions. We submitted the selected questions to VAs in a 3×3×6 fractional factorial design experiment with 3 developers (ie, Amazon, Apple, and Google), 3 modalities (ie, voice only, voice and display, display only), and 6 diseases. We assessed the rate of error-free voice responses and classified the web sources based on previous research (ie, expert, commercial, crowdsourced, or not stated).

Results: Google showed the highest total response rate, followed by Amazon and Apple. Moreover, although Amazon and Apple showed a comparable response rate in both voice-and-display and voice-only modalities, Google showed a slightly higher response rate in voice only. The same pattern was observed for the rate of expert sources. When considering the response and expert source rate across diseases, we observed that although Google remained comparable, with a slight advantage for LCA and CKD, both Amazon and Apple showed the highest response rate for LCA. However, both Google and Apple showed most often expert sources for CVA, while Amazon did so for DM.

Conclusions: Google showed the highest response rate and the highest rate of expert sources, leading to the conclusion that Google Assistant would be the most reliable tool in responding to questions about NCD management. However, the rate of expert

sources differed across diseases. We urge health organizations to collaborate with Google, Amazon, and Apple to allow their VAs to consistently provide reliable answers to health-related questions on NCD management across the different diseases.

(J Med Internet Res 2021;23(12):e32161) doi:10.2196/32161

KEYWORDS

voice assistants; conversational agents; health literacy; noncommunicable diseases; mobile phone; smart speaker; smart display; evaluation; protocol; assistant; agent; literacy; audio; health information; management; factorial; information source

Introduction

Background

Noncommunicable diseases (NCDs) constitute a significant burden on public health [1,2] and are best controlled through self-management practice [3,4]. For this purpose, digital health technologies have been developed to assist individuals in managing disease in a scalable way and at a low cost [5]. Among other types of support [6], digital health tools can facilitate self-information on an on-demand basis (eg, decision support or information lookup devices).

Voice Assistants for Information Lookup

Commercial voice assistants (VAs) can be employed on both smartphones and smart speakers, which are increasingly present in our daily lives. In fact, in 2018, 76% and 45% of individuals in advanced and emerging economies, respectively, owned a smartphone [7], and 66.4 million Americans owned a smart speaker [8]. Furthermore, although 148 million US adults used voice search for general purposes [9], 19.1 million used VAs for health-related purposes, such as gathering information about symptoms, medication, treatment options, or healthcare facilities [10]. Thus, VAs are not only scalable but also show considerable penetration.

In addition, VAs leverage speech-based interaction, which makes information lookup more efficient and accessible compared to classical methods (eg, desktop web search [11,12]). This is particularly the case in situations in which users have their hands occupied [13-17], lack reading and writing skills [18], or suffer from visual, motor, or cognitive inabilities [19-24] and cannot access information on display devices, such as smartphones, desktops, or tablets. Hence, VAs are a powerful alternative to provide chronic patients with easy access to information about NCD management. The question remains how well they can respond to health-related questions to facilitate NCD management.

Related Work

Previous research investigated VAs' reliability and assessed their ability to provide patients with reliable responses to health-related questions. These included mental and physical health, interpersonal violence [25], sexual health [26], smoking cessation [27], general health and lifestyle prompts [28], vaccines [29], addiction [30], postpartum depression [31], and COVID-19 [32]. The assessment methods varied across studies. Reliability was evaluated in terms of the ability to help with a safety-critical situation [25,28], the information correctness in comparison to official sources [31], the propensity to direct the user to available treatments or treatment referral services upon help seeking [30], or the reliability of the sources behind the responses [26,27,29,32]. All these studies have reported a rather poor performance, which is not surprising as this phenomenon has also been observed in other domains, such as shopping-related questions [9]. However, it is difficult to conclude which VA is the most reliable. For instance, although Google Assistant seems to be more reliable than Apple Siri for smoking cessation information delivery [27], the performance reverses for general lifestyle prompts [28]. Moreover, Google Assistant and Apple Siri are both more reliable than Amazon Alexa for vaccine-related questions [29]. Thus, it is unclear which VA best supports patients with chronic medical conditions in accessing information about NCD management.

Objectives

This study evaluates the ability of common Vas, such as Amazon Alexa, Apple Siri, and Google Assistant, to vocally respond to questions related to NCD management, based on expert web sources. In particular, we measure reliability through *response rate* and *source type*. Based on previous research, we control for the effect of *developer* (ie, Amazon, Apple, and Google) and interaction *modality* (ie, voice-based interaction and multimodal interaction) [28] and compare the measures to web search results [27]. This comparison allows us to relativize the reliability of VAs to the standard consumer-accessible method of information retrieval. Conducting a web search may not always be the best solution, compared to consulting medical professionals [33], but patients are increasingly searching the internet for information [12]. Therefore, we consider web search as the method of reference for health information lookup.

Hence, we seek to answer the following research questions:

1. Is the *response rate* dependent on *developers* and *modality*?
2. Is the *source type* dependent on *developers* and *modality*?
3. Is the *response rate* dependent on *developers* and *disease*?
4. Is the *source type* dependent on *developers* and *disease*?

Methods

Our experiment consisted of a quality assurance evaluation, where experimenters submitted validated questions to VAs and assessed (1) whether an error-free voice response was provided and (2) the category of the referenced web source.

Research Design

We manipulated 3 independent variables: interaction *modality*, software *developer*, and *NCD*. We set a 3×3×6 fractional design with 3 *modalities* (ie, *voice only*, *voice and display*, *display only*), 3 *developers* (ie, *Amazon*, *Apple*, and *Google*), and 6 *diseases* (ie, *lung cancer [LCA]*, *chronic obstructive pulmonary disease [COPD]*, *diabetes mellitus [DM]*, *cardiovascular disease [CVD]*, *chronic kidney disease [CKD]*, *and cerebrovascular accident [CVA]*). For the *display-only* modality, we solely included the developer *Google*. The dependent variables were the *response rate* (ie, percentage of provided voice responses) and the *source type* (ie, *expert*, *commercial*, *crowdsourced*, or *not stated*).

Measures

To measure the reliability of the different VAs, we assessed the *response rate* and *source type*. To assess the response rate, we computed the percentage of error-free speech-delivered responses. Specifically, a response was included if (1) the VA did not manifest an error, such as a system error (ie, a bug in the execution of a command), a natural language processing (NLP) error (ie, misunderstanding of the user's utterance), or an intent error (ie, the user's utterance is understood but leads to an unsupported command or an inappropriate execution), and (2) the response was voice-delivered without prompting the user to access information by interacting with a display device. For instance, if a VA was to reply *Here is what I found* and show results on the screen, we considered this response as not provided. Note that we leverage the use of VA for accessible, hands-free health-related information provision. Thus, we intentionally considered only those responses that could benefit individuals who cannot interact with a display.

In the *display-only* condition, we calculated the *response rate* by including all responses provided in the form of a web search featured snippet. A featured snippet is a unique box presented above the list of Google's search results, containing a text answer to a question-like search query (what, how, when, etc). A snippet contains a title, a summary of the answer, and a link to the web source [34]. We chose this method to compare the ability of VAs to utter a selected response with that of the web search in pulling the best information source answering the question.

The source type was categorized based on the work of Alagha and Helbing [29] and Boyd and Wilson [27] into *expert*, *commercial*, *crowdsourced*, or *not stated*. Table 1 provides a description and examples of these categories. In particular, Alagha and Helbing [29] evaluated the VAs' responses with points and categorized the web sources as *government*, *heath nonprofit*, *commercial*, *crowdsourced*, or *not stated*. If a web source was categorized as *government* and *heath nonprofit* (which they defined as *expert sources*), the response would gain a point and would otherwise get no points. Following this approach, we merged *government* and *heath nonprofit* categories into one that we called *expert* and considered such sources as *reliable* because they consistently assure good quality of information. Furthermore, we verified the reliability by exploring primary websites or by finding third-party sources of reliability evaluation (eg, Verywell Health is described as verified by doctors and collaborating with Cleveland Clinic, or Cancer.net presents patient information from the American Society of Clinical Oncology).

Table 1. Description of source types with examples.

Source type	Description	Examples
Expert	Site from a health organization representing a group of medical professionals, a governmental medical department or agency, a nonprofit medical organization, or a medical journal. These are considered reliable because they consistently provide verified and impartial information.	US Centers for Disease Control and Prevention (CDC), Mayo Clinic, *British Journal of Cancer*
Commercial	Any not nonprofit site that publishes medical information. These may or may not provide verified and impartial information.	WebMD, Medscape
Crowdsourced	Information from a site that is based on the collaboration of a large group of people. This information may or may not be verified and impartial.	Wikipedia, Blurtit
Not stated	The source type was not stated explicitly.	—[a]

[a]Not available.

Selection of Health-Related Questions

Following Kvedar et al [5], we aimed to cover NCDs involving conditions considered the leading causes of mortality. Hence, we selected questions for Alzheimer's disease (AD), LCA, COPD, DM (type 1 and type 2), CVD, CKD, and CVA. Regarding cancer, we decided on LCA as it is the most prevalent type of cancer affecting both sexes [35].

The questions were selected in 2 steps. First, we generated questions by collecting frequently asked questions (FAQs) from health organization, government, medical nonprofit, and other recognized health-related websites (eg, mayoclinic.org, nia.nih.gov, everydayhealth.com, webmd.com). When relevant, we looked for or replicated prevalent general questions across diseases. For instance, as we found the question *IsAlzheimer's disease genetic?* on nia.nih.gov and *Is lung cancer genetic?* on foxchase.org, we searched for equivalents (eg, *Are strokes genetic?* on saebo.com) or arbitrarily created one for other diseases (eg, *Is chronic kidney disease genetic?*). This process allowed us to obtain a more comprehensive pool of questions.

Second, we asked practicing medical specialists to think about their medical consultations with patients suffering from the relative type of chronic disease and to rate the frequency of occurrence of the generated field-related question on a 5-point

Likert scale (ie, 1=never, 2=rarely, 3=sometimes, 4=often, and 5=always). If they deemed it necessary, the specialists were also free to add manually missing frequent questions. For each disease, we recruited 2 specialists, resulting in a total set of 14 evaluations. We intentionally let the medical specialists rate questions in terms of absolute frequency and not ask, for instance, to indicate the *n most frequent questions*, as we wanted to avoid selection biases. The evaluations led us to an insufficient number of questions for AD. This can be explained by the fact that we asked the specialists to select questions that patients would ask their physician, and in the case of AD, information exchange rather happens between the physician and the caregiver, while the patient tends to be less involved [36]. Such tendency was confirmed by unsolicited comments from the specialists. Thus, following an intense discussion among the coauthors, we excluded the questions related to AD.

Next, we included the 10 most frequent questions for all other diseases. If any resulted in a different formulation of the same question, only the question with the simplest formulation was included (eg, we favored *How long can I live with COPD?* over *What is the life expectancy for a COPD patient?*). If the ratings did not result in a clean cut of 10 questions, we defined further selection steps. If there were more than 10 questions, we favored wh-questions (ie, questions requiring an informative answer, rather than yes or no) and removed the most ambiguous questions (eg, we excluded *Why do I have difficulties breathing?* for COPD). If there were less than 10 questions, we included questions with lower ratings, always following the criteria mentioned above. Table 2 shows the number of questions included before and after the validation process (see the complete list of selected questions in Multimedia Appendix 1).

Table 2. Number of questions before validation from medical specialists.

NCD[a]	Questions generated (N=607), n (%)
AD[b]	48 (7.9)
LCA[c]	145 (23.9)
COPD[d]	60 (9.9)
DM[e]	135 (22.2)
CVD[f]	93 (15.3)
CKD[g]	69 (11.4)
CVA[h]	57 (9.4)

[a]NCD: noncommunicable disease.

[b]AD: Alzheimer's disease.

[c]LCA: lung cancer.

[d]COPD: chronic obstructive pulmonary disease.

[e]DM: diabetes mellitus.

[f]CVD: cardiovascular disease.

[g]CKD: chronic kidney disease.

[h]CVA: cerebrovascular accident.

Setting and Apparatus

The experiment took place in a meeting room at ETH Zurich, Switzerland, between February 18 and March 4, 2021. Each VA device was placed on a table and tested separately. Two experimenters sat at the two ends of the table. One experimenter submitted the questions to each device through a text-to-speech conversion program [37] on a laptop (Lenovo ThinkPad X1 Carbon Gen 8, Lenovo Group Limited). After pilot testing, we decided to use a female voice with a speech rate of 150 words per minute. For the questions to be adequately detected by the VAs, we played the questions through a set of 2 speakers (Sony Vaio VGP-SP1, VAIO Corporation) at a distance of ca. 10 cm from the VA device.

The other experimenter took note of the voice responses and the web source through either accessing them on the user accounts (smart speakers) or taking screenshots (display devices). To have a backup of the voice responses, we used an audio recorder (Philips DVT4010, Koninklijke Philips N.V.).

Tested VAs

Based on Kocaballi et al [28], we tested commonly used unimodal and multimodal VAs. To operationalize the variables *developer* and *modality*, we employed the 3 most common Vas (ie, Amazon Alexa, Apple Siri, and Google Assistant) and aimed for the 2 most frequently used devices (ie, smart speaker and smartphone) for each VA [10]. In particular, we used Amazon Echo Dot, Apple HomePod Mini, and Google Nest Mini for the *voice-only* modality and a mix of smart displays (Amazon Echo Show) and smartphones (Apple iPhone 8 with iOS 14.4, Nokia 6.1 with Android 9) for the *voice-and-display* modality. The latter heterogeneity in the devices was due to the unavailability of the Amazon Alexa smartphone application in the authors' country of affiliation at the time of testing.

Moreover, to operationalize the *display-only* modality and our reference method of (health) information retrieval, we included a laptop with the Google Search engine (Lenovo ThinkPad X1

Carbon Gen 6 with Google Chrome). We referred to this condition as the *web search*.

All devices were set to factory settings, and we used new dedicated accounts, whose history was deleted before testing each device. Table 3 summarizes the implementation of the research design.

Table 3. Operationalization of the independent variables.

Modality	Amazon	Apple	Google
Voice only	Smart speaker (Eco Dot)	Smart speaker (HomePod Mini)	Smart speaker (Nest Mini)
Voice and display	Smart display (Eco Show)	Smartphone (iPhone 8, iOS 14.4)	Smartphone (Nokia 6.1, Android 9)
Display only	—[a]	—	Laptop (Lenovo ThinkPad X1 Carbon Gen 6, Google Chrome)

[a]Not available.

Procedure

The selected questions were played in randomized order. In the *voice-only* and *voice-and-display* modalities, upon manual start, the text-to-speech program would play the appropriate wake-up keyword (ie, “Hey Alexa,” “Hey Siri,” or “Hey Google”), followed by a question. In the case of an error (see the Measures section), we first replayed the question and then, if necessary, played the question again using a male voice. In the case of a persistent error, 1 of the experimenters asked the question manually. This protocol allowed us to ensure the ability to respond to a question did not depend on the input quality. If none of those attempts produced a voice response, we considered the response as not provided. In the *display-only* modality, the question was directly entered in the text field of the web search engine.

Ethics

Given the involvement of practicing medical specialists, we validated our research proposal (EK 2020-N-173) with the ETH Zurich Ethics Commission. The procedure was approved without reservation on December 21, 2020.

Statistical Analysis

R Version 1.2 (RStudio, Inc.) was used to compute the frequency and descriptive statistics.

For the sake of comparison, all results (response rate and source type) are shown in percentages.

Results

For each subsection, we first introduce the analysis and then present the results in the form of a table (see also Multimedia Appendix 2 for the complete list of voice responses, web sources, and categorization).

Response Rate Across Developers and Modalities

To understand to what extent the examined VAs could provide an answer at all, we calculated the response rate across developers and modalities. The results are summarized in Table 4.

Table 4. Response rate across developers and modalities.

Response	Display only (N=60), n (%)	Voice and display (N=180), n (%)	Voice only (N=180), n (%)	Total (N=420), n (%)
Amazon				
No	—[a]	14 (23.3)	15 (25)	29 (24.2)
Yes	—	46 (76.7)	45 (75)	91 (75.8)
Total	—	60 (100)	60 (100)	120 (100)
Apple				
No	—	48 (80)	47 (78.3)	95 (79.2)
Yes	—	12 (20)	13 (21.7)	25 (20.8)
Total	—	60 (100)	60 (100)	120 (100)
Google				
No	—	6 (10)	0	6 (5)
Yes	—	54 (90)	60 (100)	114 (95)
Total	—	60 (100)	60 (100)	120 (100)
Web search				
No	12 (20)	—	—	12 (20)
Yes	48 (80)	—	—	48 (80)
Total	60 (100)	—	—	60 (100)

[a]Not available.

Source Type Across Developers and Modalities

Classifying the web sources into *commercial*, *crowdsourced*, *expert*, and *not stated* allowed us to derive the level of reliability of the voice response across developers and modalities. The results are summarized in Table 5.

Table 5. Source type across developers and modalities.

Source type	Display only (N=48), n (%)	Voice and display (N=112), n (%)	Voice only (N=118), n (%)	Total (N=278), n (%)
Amazon				
Commercial	__[a]	11 (23.9)	9 (20)	20 (22)
Crowdsourced	—	4 (8.7)	4 (8.9)	8 (8.8)
Expert	—	27 (58.7)	28 (62.2)	55 (60.4)
Not stated	—	4 (8.7)	4 (8.9)	8 (8.8)
Total	—	46 (100)	45 (100)	91 (100)
Apple				
Commercial	—	2 (16.7)	2 (15.4)	4 (16)
Crowdsourced	—	6 (50)	3 (23.1)	9 (36)
Expert	—	1 (8.3)	1 (7.7)	2 (8)
Not stated	—	3 (25)	7 (53.8)	10 (40)
Total	—	12 (100)	13 (100)	25 (100)
Google				
Commercial	—	12 (22.2)	12 (20)	24 (21.1)
Crowdsourced	—	3 (5.6)	3 (5)	6 (5.3)
Expert	—	39 (72.2)	45 (75)	84 (73.7)
Total	—	54 (100)	60 (100)	114 (100)
Web search				
Commercial	15 (31.3)	—	—	15 (31.3)
Expert	33 (68.8)	—	—	33 (68.8)
Total	48 (100)	—	—	48 (100)

[a]Not available.

Furthermore, we want to point out that all error-free responses involved the synthesis of a meaningful response and none resulted in the VA proposing to use a voice application to answer the question (eg, Alexa Skill or Google Action).

Response Rate Across Developers and Diseases

We aimed to verify whether there was a pattern in the ability to provide an answer depending on the NCD in question. As the effect of modality was minimal, we present the results for the *voice-only* modality. Thus, we calculated the percentage of provided answers across *developer* and *disease*. Table 6 summarizes the results.

Table 6. Response rate by developer and disease.

Response	LCA[a] (N=70), n (%)	COPD[b] (N=70), n (%)	DM[c] (N=70), n (%)	CVD[d] (N=70), n (%)	CKD[e] (N=70), n (%)	CVA[f] (N=70), n (%)	Total (N=420), n (%)
Amazon							
No	—[g]	8 (40)	2 (10)	6 (30)	7 (35)	6 (30)	29 (24.2)
Yes	20 (100)	12 (60)	18 (90)	14 (70)	13 (65)	14 (70)	91 (75.8)
Total	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	120 (100)
Apple							
No	8 (40)	16 (80)	19 (95)	18 (90)	18 (90)	16 (80)	95 (79.2)
Yes	12 (60)	4 (20)	1 (5)	2 (10)	2 (10)	4 (20)	25 (20.8)
Total	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	120 (100)
Google							
No	—	1 (5)	1 (5)	2 (10)	—	2 (10)	6 (5)
Yes	20 (100)	19 (95)	19 (95)	18 (90)	20 (100)	18 (90)	114 (95)
Total	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	20 (100)	120 (100)
Web search							
No	1 (10)	1 (10)	1 (5)	3 (30)	5 (50)	1 (10)	12 (20)
Yes	9 (90)	9 (90)	9 (45)	7 (70)	5 (50)	9 (90)	48 (80)
Total	10 (100)	10 (100)	10 (100)	10 (100)	10 (100)	10 (100)	60 (100)

[a]LCA: lung cancer.

[b]COPD: chronic obstructive pulmonary disease.

[c]DM: diabetes mellitus.

[d]CVD: cardiovascular disease.

[e]CKD: chronic kidney disease.

[f]CVA: cerebrovascular accident.

[g]Not available.

Source Type Across Developers and Diseases

Calculating the proportion of source types across developers allowed assessing the presence of information reliability patterns among the VAs. We present the results for the *voice-only* modality. Table 7 summarizes our results.

Table 7. Proportion of source type by developer and disease.

Source type	LCA[a] (N=70), n (%)	COPD[b] (N=70), n (%)	DM[c] (N=70), n (%)	CVD[d] (N=70), n (%)	CKD[e] (N=70), n (%)	CVA[f] (N=70), n (%)	Total (N=420), n (%)
Amazon							
Commercial	2 (10)	4 (33.3)	6 (33.3)	3 (21.4)	3 (23.1)	2 (14.3)	20 (22)
Crowdsourced	4 (20)	0	0	0	2 (15.4)	2 (14.3)	8 (8.8)
Expert	10 (50)	6 (50)	12 (66.7)	9 (64.3)	8 (61.5)	10 (71.4)	55 (60.4)
Not stated	4 (20)	2 (16.7)	0	2 (14.3)	0	0	8 (8.8)
Total	20 (100)	12 (100)	18 (100)	14 (100)	13 (100)	14 (100)	91 (100)
Apple							
Commercial	4 (33.3)	0	0	0	0	0	4 (16)
Crowdsourced	5 (41.7)	0	0	1 (50)	1 (50)	2 (50)	9 (36)
Expert	0	0	0	0	0	2 (50)	2 (8)
Not stated	3 (25)	4 (100)	1 (100)	1 (50)	1 (50)	0	10 (40)
Total	12 (100)	4 (100)	1 (100)	2 (100)	2 (100)	4 (100)	25 (100)
Google							
Commercial	1 (5)	10 (52.6)	5 (26.3)	4 (22.2)	3 (15)	1 (5.6)	24 (21.1)
Crowdsourced	4 (20)	0	0	0	2 (10)	0	6 (5.3)
Expert	15 (75)	9 (47.4)	14 (73.7)	14 (77.8)	15 (75)	17 (94.4)	84 (73.7)
Total	20 (100)	19 (100)	19 (100)	18 (100)	20 (100)	18 (100)	114 (100)
Web search							
Commercial	2 (22.2)	5 (55.6)	4 (44.4)	1 (14.3)	2 (40)	1 (11.1)	15 (31.3)
Expert	7 (77.8)	4 (44.4)	5 (55.6)	6 (85.7)	3 (60)	8 (88.9)	33 (68.8)
Total	9 (100)	9 (100)	9 (100)	7 (100)	5 (100)	9 (100)	48 (100)

[a]LCA: lung cancer.

[b]COPD: chronic obstructive pulmonary disease.

[c]DM: diabetes mellitus.

[d]CVD: cardiovascular disease.

[e]CKD: chronic kidney disease.

[f]CVA: cerebrovascular accident.

Discussion

Principal Results

Google showed the highest response rate and rate of expert sources, leading to the conclusion that Google Assistant would be the most reliable tool in responding to questions about NCD management. However, the rate of expert sources differed across diseases.

Response Rate Across Developers and Modalities

We observed Google Assistant provided the highest response rate, even outperforming the web search results. Apple Siri showed the lowest response rate. This specific advantage of Google Assistant is consistent with previous studies [27,29,32,38,39].

Moreover, Apple Siri often replied *I found this on the web* and presented visually a list of results instead of *voicing* a unique response. This may reflect a tendency to transfer the responsibility of information retrieval to the patients, whereas they are in charge of choosing (the most) reliable information source. As we believe in the potential of VAs in increasing the accessibility of health information by *vocally* interacting with patients having their hands occupied or with disabilities [13-17,19-24], we urge Apple to consider favoring voice responses over lists of results to make the information search more suitable for hands-free interaction.

Source Type Across Developers and Modalities

Similar to the results for the response rate, Google Assistant also answered the most frequently with *expert* sources, without outperforming the web search results. Amazon Alexa showed 13.3% fewer *expert* sources. These results are surprising, given the partnership that Amazon started with the United Kingdom National Health Service (NHS) in 2019, allowing the former to freely access health-related information provided by the latter

[40]. Finally, Apple Siri showed the least *expert* sources, favoring *crowdsourced* sources. Importantly, Apple Siri also showed the highest proportion of missing sources (ie, *not stated*). Alagha and Helbing [29] classified *commercial*, *crowdsourced*, and *not stated* sources as less reliable than *expert* sources. As *commercial* and *crowdsourced* sources may, in some cases, still provide reliable information [41,42], Apple Siri may tend to transfer the responsibility of judging the reliability of the information to the patient. This assumption is also supported by the low rate of voice responses from Apple Siri discussed above.

Response Rate Across Developers and Diseases

When comparing the response rate across diseases, we observed Google Assistant to respond rather similarly across diseases, with a slight advantage for LCA and CKD. Amazon Alexa and Apple Siri showed the highest response rate for LCA. Thus, questions related to LCA seemed to have a general advantage over the other diseases. Both Amazon Alexa and Google Assistant outperformed web search results on both LCA and CKD.

Source Type Across Developers and Diseases

Our results showed Google Assistant to most often use *expert* sources for CVA, while Amazon Alexa did so for DM. Apple Siri showed *expert* sources only for questions about CVA. Thus, there seems to be an advantage of CVA questions in being reliable the most often, except when those questions are asked to Amazon Alexa.

Finally, this slight heterogeneity in results is in line with the diversity in previous research, whereas depending on the condition or disease of interest, reliability varies across VAs [27-29]. Hence, there seems to be a need for systematization of health information search algorithms across the different medical domains.

Limitations

Despite our best efforts, our study presents some limitations.

First, for technical reasons, we employed a smart display instead of a smartphone for Amazon Alexa. The observed similarity in source type proportion between modalities may be explained by the use of 2 types of home devices. This similarity may reflect a consistency in information provision across Amazon Alexa modalities, which is desirable. However, future research should aim to replicate the results by comparing the reliability between the Amazon Alexa app and an Amazon smart speaker. This will support an absence of the effect of interaction modality in Amazon Alexa.

Second, to control for the effect of time on the responses, we aimed to restrict the time window inside which to test the VAs as much as possible. Thus, given the high number of questions submitted (ie, 60 questions per device, resulting in a total of 420 submissions), we did not submit the same questions multiple times. However, we conducted a post hoc test-retest reliability assessment by randomly selecting 1 question per NCD and submitting each one 10 times to all VAs. Our results showed no variation in the voice responses or the source type. Nevertheless, as observed in previous research that VAs do not always provide the same response [29], future research should consider the test-retest reliability of VAs in responding to a more extensive set of questions about the included NCDs.

Third, questions were selected by looking for FAQ pages on health-related websites, but it is difficult to conclude whether all relevant questions were included. As we shared the list of questions (see Multimedia Appendix 1), we hope future research will be able to establish whether additional relevant questions need to be tested.

Fourth, the source type was categorized based on the work of Alagha and Helbing [29] and Boyd and Wilson [27] into *expert*, *commercial*, *crowdsourced*, or *not stated*, whereas *expert* sources represented *government* and *heath nonprofit* sources [29]. We considered such sources as the *mostreliable* because they consistently assured good quality of information. However, although not stating a source of information makes it difficult for a patient to judge the response's trustworthiness, *commercial* and *crowdsourced* sources may, in some cases, still provide correct information. Health-related information coming, for instance, from wikipedia.com (*crowdsourced*) varies importantly in terms of quality [41] and could, in some cases, still provide reliable information. *Commercial* sources may also contain partly reliable information [42], despite presenting a higher risk of bias toward marketing purposes [43]. Future research should investigate directly the reliability of the provided information rather than the mere source type in order to have a more fine-grained landscape of its reliability.

Fifth, the FAQs about AD were evaluated as rather rarely asked by the patients. As information exchange about AD rather happens between the physician and the caregiver, while the patient tends to be less involved [36], future research should include questions from caregivers as well in order to assess the response and *expert* source rates for FAQs related to this disease.

Finally, questions were validated by Swiss doctors and thus may not be representative of the patient's most frequent concerns in other realities. Future research should validate the tested questions with professionals of other countries to ensure their relevance across nations.

Comparison With Related Work

Comparing VAs to Web Search

Our results are partly in line with the work of Boyd and Wilson [27], comparing the ability of Apple Siri and Google Assistant (*voice-and-display* modality only) to provide expert information for smoking-cessation-related questions compared to the web search. Similar to Boyd and Wilson [27], we observed Google Assistant to be more reliable than Apple Siri at providing information from reliable sources (which in their study was defined as web pages of *health agencies with medical expertise*). However, although Boyd and Wilson [27] found the web search to be more reliable better than Google Assistant, our evaluation showed Google Assistant to be slightly more reliable than the web search (in both *voice-and-display* and *voice-only* modalities). The difference may lie in the evaluation of the web search's responses: although we considered featured snippets as the selected response to judge the search engine on, Boyd and Wilson [27] considered the *first non-advertisement link or*

information of the list of results. This criterion may have led the authors to collect more responses from the web search and thus a different distribution of reliable sources. Moreover, Boyd and Wilson [27] were the only ones comparing the reliability of the VAs to a traditional method of health information search, such as browser-based web search. Not comparing the VAs' response rate and information reliability to web search makes it difficult to conclude on their absolute ability to respond to health-related questions. Our results not only show Google Assistant leading in NCD-related information provision, but also that it can even outperform the web search results and quite successfully inform patients about NCD management.

Evaluating Response Source Type

Based on Alagha and Helbing [29], who based their evaluation system on Boyd and Wilson [27], we evaluated the source type and classified the sources as *not stated*, *crowdsourced*, *commercial*, or *expert* (ie, a combination of the *health nonprofit* and *government* sources in Alagha and Helbing [29]). Although the authors observed Apple Siri and Google Assistant responding and providing expert information more frequently than Amazon Alexa (in the *voice-and-display* modality only), our study showed an advantage of Google Assistant, followed by Amazon Alexa. The higher advantage of Amazon Alexa observed in our study may be explained by the time of data collection. More specifically, Alagha and Helbing [29] tested the VAs in 2018, which was before Amazon would instantiate a partnership with the NHS in 2019 to provide reliable health information [40] (see also the Principal Results section). Thus, Amazon Alexa's ability to respond to health-related questions and the reliability of its sources may have increased since then.

Response Reliability Versus Response Appropriateness

Considering our results and the work of Boyd and Wilson [27] and of Alagha and Helbing [29], Google Assistant seems to be the best solution for health-related information lookup and thus for best supporting patients with NCDs. This conclusion may, however, be challenged by studies by Kocaballi et al [28] and Yang et al [31].

Kocaballi et al [28] assessed how frequently Amazon Alexa, Apple Siri, and Google Assistant would provide *appropriate* responses to safety-critical and non-safety-critical questions. Appropriateness was defined as the VA *recommending to get help from a health professional or service and to provide specific contact information* if the question was safety-critical and as *including relevant information to solve the problem* raised in the question if the question was non-safety-critical [28]. The author observed that although Google Assistant often provided a web source with its response, Apple Siri was the one providing the highest number of appropriate responses. Similarly, Yang et al [31], who assessed the *clinical appropriateness* of VAs' responses, observed Google Assistant to provide an *appropriate* response only 21% of the time, while Amazon Alexa performed slightly better, with 29% of *appropriate* responses. *Appropriateness* was defined by 2 physicians comparing the response provided by the VA to the answer present in the American College of Obstetricians and Gynecologists patient-focused FAQs. Although *appropriateness* evaluation is more meticulous because of analyzing the content, it is more subjective. We approached the responses from a *reliability* perspective and assessed them solely on an objective level, that is, by assessing the use of recognized health web sources. Given that, as mentioned above, source type may not be sufficient to evaluate information quality, and future research should combine source evaluation with professional content evaluation to obtain a more complete representation of the VAs' ability to provide patients with reliable information.

Leveraging Speech Interaction

In general, most of the related work presented above [27-29] evaluated VAs' responses by considering both display and voice responses. That is, if the VA was not to vocally synthesize a direct response to the question but to say *Here's what I found* and visually showed a list of results, the first of that list was still considered for evaluation. In our study, we considered only voice responses. The rationale behind this decision is that the use of VAs for (health-related) information lookup is truly advantageous and accessible if it can breach the barriers of lack of literacy [18]; visual, motor, or cognitive inabilities [19-24]; or manual unavailability [13-17]. Only Yang et al [31] reported whether the VAs provided a voice response. Although the differences between VAs were not statistically significant, the authors showed Apple Siri to be the least reliable and Amazon Alexa to be the most reliable. Our results replicate the low voice response rate of Apple Siri but show Google Assistant to respond vocally more frequently than Alexa. Given the small sample of questions and low statistical power in Yang et al [31], future research should evaluate a larger sample and contrast the results to the findings of Yang et al [31] and this study.

FAQ Relevance

Although the questions used in the studies discussed above were based on health-related web pages [27,29,31], none of them verified the questions' relevance for the patients themselves in the real-world practice. In particular, gathering questions from health-related websites ensured including questions that are of most interest to the affected population, that is, not only patients but potentially also their caregivers and close social network. However, we aimed to target questions specifically relevant to the patients, as they are the protagonists of self-management. Thus, we submitted the selected questions to the respective practicing medical specialists and explicitly asked them to rate their frequency, considering the questions coming from the patients. The fact that not all questions were relevant is also supported by the fact that specialists did not select all questions as being "Often" to "Always" asked by patients (see also Multimedia Appendix 1).

Implications

Complications related to NCDs, such as CVD, cancer, chronic respiratory disease, and DM, are among the main causes of mortality, accounting for 71% and 74% of all deaths worldwide in 2016 [1] and 2019 [2], respectively. Preventing those complications or their worsening is, therefore, crucial for survival. Engaging with self-management solutions is the best preventive practice [3,4]. VAs can support self-management through efficient and accessible information delivery by fostering patient's health literacy [44,45] (ie, the ability to

obtain, process, understand, and communicate health-related information to a level that favors positive health behavior [46]). For instance, an individual with a chronic respiratory disease who can ask their VA about smoking cessation strategies and put those into practice has higher chances to stop smoking and mitigate their health conditions. Our results show that Amazon Alexa and Google Assistant are capable of providing reliable health information through pure speech-based interaction, although such ability differs across NCDs.

Conclusion

This study provides evidence of the ability of Vas, such as Amazon Alexa, Apple Siri, and Google Assistant, to provide reliable voice responses to questions related to NCD management compared to the standard consumer-accessible method of information lookup (ie, web search). We validated NCD-related questions with practicing medical professionals, submitted a set of 60 questions to each VA, and assessed the response rate and source type. We answered the first research question (ie, *Is the response rate dependent on developers and modality?*), observing that Google Assistant responded to most of the answers and Apple Siri responded to the fewest. Modality played a minimal role, whereas Google Assistant and Apple Siri responded slightly more often in the *voice-only* modality and Amazon Alexa in the *voice-and-display* modality. Moreover, we answered our second research question (ie, *Is the source type dependent on developers and modality?*), finding that Google Assistant based most of its responses on *expert* sources of information, even outperforming the web search snippets. Furthermore, Amazon Alexa was less reliable but provided *expert* sources more than 50% of the times. Finally, Apple Siri was the least reliable, providing a considerable percentage of *crowdsourced* sources or often not providing a source at all. Across modalities, Amazon Alexa and Apple Siri similarly provided *expert* sources, while Google Assistant did so more frequently in the *voice-only* modality. Thus, the variation seemed to be more influenced by *developer* than *modality*. Finally, answering our third and fourth research questions (ie, *Is the response rate dependent on developers and disease?* and *Is the source type dependent on developers and disease?*), we observed that although there is a slight variation across the diseases, Google Assistant showed a general clear advantage. Nevertheless, although Google Assistant seems to be a good option to ask NCD-related questions, a large number of *commercial* sources was used, in particular for COPD. Providing patients with unverified or non-evidence-based information may be counterproductive if not dangerous. Therefore, we call out health organizations to collaborate with technology companies, such as Google, Amazon, and Apple, to ensure patients with NCDs are provided with openly reliable (ie, expert) information about the management of their condition. Finally, as the algorithms behind the VAs continuously change, future research should establish the temporal consistency of these results.

To conclude, our contributions lie in the following aspects. First, to the best of our knowledge, no previous research assessed the reliability of the 3 most prevalent VAs in responding to NCD-related questions. Second, we tested the selected VAs by submitting questions that we validated with practicing medical specialists in terms of the frequency of occurrence in their medical consultations. Third, we systematically controlled for the effect of visual display by testing both smart speakers and display devices for each VA. Finally, as previous research remains rather preliminary and lacks transparent method reporting [47], we aimed to report our methods as precisely as possible in the hope of stimulating informed future research on the ability of VAs to retrieve reliable information about health-related topics.

Acknowledgments

We would like to thank Dr med. Christoph Globas, Dr med. Dan Georgescu, Dr med. Daniel Franzen, Dr med. Eva Krebs-Roubicek, Dr med. Eva Bühlmann, Dr med. Frank Rassouli, Dr med. Martina Pechula Thut, Dr med. Michael Kiessling, Dr med. Oliver Gämperli, Dr med. Rebekka Gigger, Dr med. Susanne Wegener, Dr med. Thomas Schachtner, Dr med. Thomas Züger, and Dr med. Viviane Hess for validating our health-related questions.

The research was conducted at Future Health Technologies at the Singapore-ETH Centre, which was established collaboratively between ETH Zurich and the National Research Foundation Singapore. This research was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) program. This work was also supported by CSS Insurance, Switzerland.

Authors' Contributions

CB, EF, and TK were responsible for the study design. CB and ZFK were responsible for the selection of the questions, data collection, and response evaluation. CB was responsible for the first draft of the manuscript. All authors were responsible for critical feedback and final revisions of the manuscript.

Conflicts of Interest

At the time of submission, CB, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also cofounders of Pathmate

Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in the study described in this paper.

Multimedia Appendix 1
Complete list of selected questions.
[PDF File (Adobe PDF File), 32 KB - jmir_v23i12e32161_app1.pdf]

Multimedia Appendix 2
Complete list of the voice assistant's responses and sources.
[PDF File (Adobe PDF File), 236 KB - jmir_v23i12e32161_app2.pdf]

References

1. World Health Organization. World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2020.
2. World Health Organization. The Top 10 Causes of Death. URL: http://web.archive.org/web/20210623104746/https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2021-12-07]
3. Mackey LM, Doody C, Werner EL, Fullen B. Self-management skills in chronic disease management. Med Decis Making 2016 Apr 06;36(6):741-759. [doi: 10.1177/0272989x16638330]
4. van der Heide I, Poureslami I, Mitic W, Shum J, Rootman I, FitzGerald JM. Health literacy in chronic disease management: a matter of interaction. J Clin Epidemiol 2018 Oct;102:134-138. [doi: 10.1016/j.jclinepi.2018.05.010] [Medline: 29793001]
5. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
6. World Health Organization. Classification of Digital Health Interventions v1.0: A Shared Language to Describe the Uses of Digital Technology for Health. Geneva: World Health Organization; 2018.
7. Rosenberg S. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. 2019. URL: https://tinyurl.com/mwp2hhu6 [accessed 2021-12-06]
8. Kinsella B. Smart Speaker Consumer Adoption Report 2019. 2019. URL: https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/ [accessed 2021-12-07]
9. Kinsella B, Mutchler A. Voice Assistant SEO Report for Brands. 2019. URL: https://voicebot.ai/wp-content/uploads/2019/07/voice_assistant_seo_report_for_brands_2019_voicebot.pdf [accessed 2021-12-07]
10. Kinsella B, Mutchler A. Voice Assistant Consumer Adoption in Healthcare. 2019. URL: https://voicebot.ai/wp-content/uploads/2019/10/voice_assistant_consumer_adoption_in_healthcare_report_voicebot.pdf [accessed 2021-12-07]
11. Beverley CA, Bath PA, Booth A. Health information needs of visually impaired people: a systematic review of the literature. Health Soc Care Community 2004 Jan;12(1):1-24. [doi: 10.1111/j.1365-2524.2004.00460.x]
12. Akerkar SM, Bichile LS. Doctor patient relationship: changing dynamics in the information age. J Postgrad Med 2004;50(2):120-122 [FREE Full text] [Medline: 15235209]
13. Large D, Burnett G, Anyasodo B, Skrypchuk L. Assessing cognitive demand during natural language interactions with a digital driving assistant. 2016 Presented at: Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications; 2016; Ann Arbor, MI, USA p. 67-74. [doi: 10.1145/3003715.3005408]
14. Simmons SM, Caird JK, Steel P. A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance. Accid Anal Prev 2017 Sep;106:31-43. [doi: 10.1016/j.aap.2017.05.013] [Medline: 28554063]
15. Young R, Zhang J. Driven to distraction? A review of speech technologies in the automobile. J Assoc Voice Interact Design 2017;2(1):1-34 [FREE Full text]
16. Militello L, Sezgin E, Huang Y, Lin S. Delivering perinatal health information via a voice interactive app (SMILE): mixed methods feasibility study. JMIR Form Res 2021 Mar 01;5(3):e18240 [FREE Full text] [doi: 10.2196/18240] [Medline: 33646136]
17. Vtyurina A, Fourney A. Exploring the role of conversational cues in guided task support with virtual assistants. NY, USA: Association for Computing Machinery; 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC Canada p. 1-7. [doi: 10.1145/3173574.3173782]
18. UNESCO Institute for Statistics. Literacy Rates Continue to Rise from One Generation to the Next. 2017. URL: http://uis.unesco.org/sites/default/files/documents/fs45-literacy-rates-continue-rise-generation-to-next-en-2017_0.pdf [accessed 2021-12-07]
19. Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, et al. Investigating the accessibility of voice assistants with impaired users: mixed methods study. J Med Internet Res 2020 Sep 25;22(9):e18431 [FREE Full text] [doi: 10.2196/18431] [Medline: 32975525]
20. Pradhan A, Mehta K, Findlater L. "Accessibility came by accident": use of voice-controlled intelligent personal assistants by people with disabilities. NY, USA: Association for Computing Machinery; 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC Canada p. 1-13. [doi: 10.1145/3173574.3174033]

21. Barata M, Galih Salman A, Faahakhododo I, Kanigoro B. Android based voice assistant for blind people. LHTN 2018 Jul 27;35(6):9-11. [doi: 10.1108/lhtn-11-2017-0083]
22. Abdolrahmani A, Storer KM, Roy ARM, Kuber R, Branham SM. Blind leading the sighted. ACM Trans Access Comput 2020 Jan 20;12(4):1-35. [doi: 10.1145/3368426]
23. Choi D, Kwak D, Cho M, Lee S. "Nobody speaks that fast!" An empirical study of speech rate in conversational agents for people with vision impairments. 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI, USA. [doi: 10.1145/3313831.3376569]
24. Friedman N, Cuadra A, Patel R, Azenkot S, Stein J, Ju W. Voice assistant strategies and opportunities for people with tetraplegia. 2019 Presented at: The 21st International ACM SIGACCESS Conference on Computers and Accessibility; October 28-30, 2019; Pittsburgh, PA, USA p. 575-577. [doi: 10.1145/3308561.3354605]
25. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
26. Wilson N, MacDonald EJ, Mansoor OD, Morgan J. In bed with Siri and Google Assistant: a comparison of sexual health advice. BMJ 2017 Dec 13;359:j5635. [doi: 10.1136/bmj.j5635] [Medline: 29237603]
27. Boyd M, Wilson N. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS One 2018;13(3):e0194811 [FREE Full text] [doi: 10.1371/journal.pone.0194811] [Medline: 29590168]
28. Kocaballi AB, Quiroz JC, Rezazadegan D, Berkovsky S, Magrabi F, Coiera E, et al. Responses of conversational agents to health and lifestyle prompts: investigation of appropriateness and presentation structures. J Med Internet Res 2020 Feb 9;22(2):e15823. [doi: 10.2196/15823]
29. Alagha EC, Helbing RR. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):e100075 [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
30. Nobles AL, Leas EC, Caputi TL, Zhu S, Strathdee SA, Ayers JW. Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. NPJ Digit Med 2020;3:11 [FREE Full text] [doi: 10.1038/s41746-019-0215-9] [Medline: 32025572]
31. Yang S, Lee J, Sezgin E, Bridge J, Lin S. Clinical advice by voice assistants on postpartum depression: cross-sectional investigation using Apple Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e24045 [FREE Full text] [doi: 10.2196/24045] [Medline: 33427680]
32. Goh A, Wong L, Yap K. Evaluation of COVID-19 information provided by digital voice assistants. Int J Digit Health 2021:3. [doi: 10.29337/ijdh.25]
33. 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]
34. Strzelecki A, Rutecka P. Featured snippets results in Google web search: an exploratory study. In: Marketing and Smart Technologies. Smart Innovation, Systems and Technologies. Singapore: Springer; 2020.
35. Mattiuzzi C, Lippi G. Current cancer epidemiology. J Epidemiol Glob Health 2019 Dec;9(4):217-222 [FREE Full text] [doi: 10.2991/jegh.k.191008.001] [Medline: 31854162]
36. Werner P, Gafni A, Kitai E. Examining physician-patient-caregiver encounters: the case of Alzheimer's disease patients and family physicians in Israel. Aging Ment Health 2004 Nov;8(6):498-504 [FREE Full text] [doi: 10.1080/13607860412331303793] [Medline: 15724831]
37. Bérubé C. MONICA. 2021. URL: https://gitlab.ethz.ch/berubec/MONICA [accessed 2021-12-07]
38. Palanica A, Thommandram A, Lee A, Li M, Fossat Y. Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. NPJ Digit Med 2019 Jun 20;2(1):55 [FREE Full text] [doi: 10.1038/s41746-019-0133-x] [Medline: 31304401]
39. Palanica A, Fossat Y. Medication name comprehension of intelligent virtual assistants: a comparison of Amazon Alexa, Google Assistant, and Apple Siri between 2019 and 2021. Front Digit Health 2021 May 19;3:669971 [FREE Full text] [doi: 10.3389/fdgth.2021.669971] [Medline: 34713143]
40. Department of Health and Social Care. News Story: NHS Health Information Available through Amazon's Alexa. 2019. URL: http://web.archive.org/web/20210706143613/https://www.gov.uk/government/news/nhs-health-information-available-through-amazon-s-alexa [accessed 2021-12-07]
41. Smith DA. Situating Wikipedia as a health information resource in various contexts: A scoping review. PLoS One 2020 Feb 18;15(2):e0228786 [FREE Full text] [doi: 10.1371/journal.pone.0228786] [Medline: 32069322]
42. Nowrouzi B, Gohar B, Nowrouzi-Kia B, Garbaczewska M, Brewster K. An examination of scope, completeness, credibility, and readability of health, medical, and nutritional information on the internet: a comparative study of Wikipedia, WebMD, and the Mayo Clinic websites. Can J Diabetes 2015 Apr;39:S71. [doi: 10.1016/j.jcjd.2015.01.267]
43. Ritchie L, Tornari C, Patel PM, Lakhani R. Glue ear: how good is the information on the World Wide Web? J Laryngol Otol 2016 Jan 25;130(2):157-161. [doi: 10.1017/s0022215115003230]

44. Xie B. Improving older adults' e-health literacy through computer training using NIH online resources. Libr Inf Sci Res 2012 Jan 01;34(1):63-71 [FREE Full text] [doi: 10.1016/j.lisr.2011.07.006] [Medline: 22639488]
45. Robinson C, Graham J. Perceived internet health literacy of HIV-positive people through the provision of a computer and internet health education intervention. Health Info Libr J 2010 Dec 1;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
46. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010 Dec;15 Suppl 2(4):9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
47. Bérubé C, Schachner T, Keller R, Fleisch E, V Wangenheim F, Barata F, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar 29;23(3):e25933 [FREE Full text] [doi: 10.2196/25933] [Medline: 33658174]

Abbreviations

AD: Alzheimer's disease
CKD: chronic kidney disease
COPD: chronic obstructive pulmonary disease
CVA: cerebrovascular accident
CVD: cardiovascular disease
DM: diabetes mellitus
LCA: lung cancer
NCD: noncommunicable disease
NHS: National Health Service
VA: voice assistant

Edited by G Eysenbach; submitted 28.07.21; peer-reviewed by B Kocaballi, E Sezgin, A Islam; comments to author 14.08.21; revised version received 19.11.21; accepted 20.11.21; published 20.12.21.

Please cite as:
Bérubé C, Kovacs ZF, Fleisch E, Kowatsch T
Reliability of Commercial Voice Assistants' Responses to Health-Related Questions in Noncommunicable Disease Management: Factorial Experiment Assessing Response Rate and Source of Information
J Med Internet Res 2021;23(12):e32161
URL: https://www.jmir.org/2021/12/e32161
doi:10.2196/32161
PMID:34932003

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org